



THE PROTEIN- PROTEIN INTERACTIONS (PPI) THAT CAUSE MACHADO JOSEPH DISEASE (SCA3)

Pedro Emanuel de Almeida Rocha

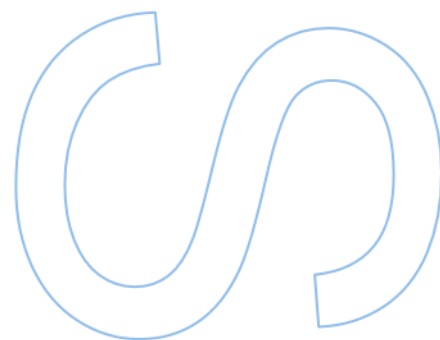
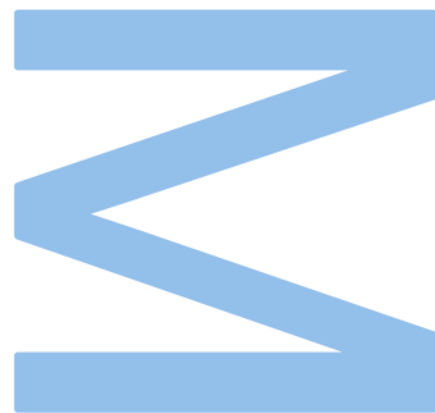
Master's degree in Cellular and Molecular Biology
Biology Department, Faculdade de Ciências da Universidade do Porto
(FCUP)
2022-2023

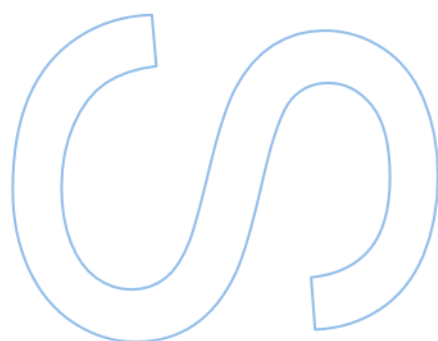
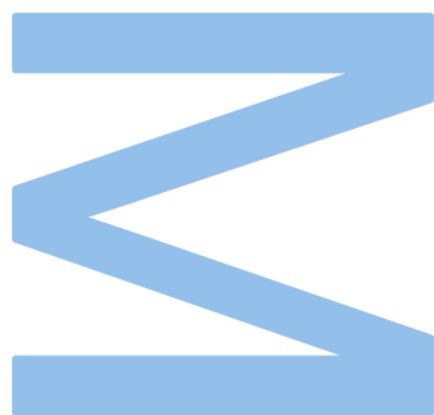
Supervisor

Doutora Cristina Vieira, Assistant Researcher, Instituto de
Investigação e Inovação em Saúde (I3S)

Co-supervisor

Doutor Jorge Vieira, Principal Researcher, I3S





Agradecimentos

Em primeiro lugar, gostaria de expressar a minha sincera gratidão à Professora Doutora Cristina Vieira, pelo convite e pela oportunidade de eu me integrar no seu projeto de investigação ao longo deste último ano. O seu apoio e orientação foram fundamentais para a realização desta tese

Queria também agradecer ao meu co-orientador Professor Doutor Jorge Vieira, pelas contribuições analíticas e pelo a assistência prestada na realização deste trabalho. Estou grato pela oportunidade de me terem integrado num ambiente de investigação fora da minha área de conforto, o que resultou em novos conhecimentos e experiências enriquecedoras.

Agradeço também a todos meus amigos que me trouxeram momentos de alegria e descontração ao longo deste período de estudos. Sem eles, a minha jornada seria muito mais monótona e solitária.

Não posso deixar de expressar a minha mais profunda gratidão aos meus pais pelo apoio dado, não apenas durante o período em que escrevi a minha tese como em toda a minha vida académica, onde foram uma fonte de apoio inabalável. Por último, mas não menos importante, ao meu “manito”. Apesar de estar sempre ocupado nas suas tarefas sempre mostrou grande preocupação e um interesse genuíno pelo meu percurso e atividades.

A todos os mencionados acima e a todos aqueles que de alguma forma contribuíram para esta conquista,

MUITO OBRIGADO

Resumo

A Ataxia Espinocerebelosa Tipo 3 (SCA3) é uma doença neurodegenerativa causada pela expansão do trato poliglutamínico (polyQ) da ataxina-3 [1]. Esta expansão da repetição trinucleotídica CAG leva à alteração estrutural da proteína codificada, o que resulta em alterações na rede nativa de interações proteína-proteína (PPIs) associadas à atividade regular da proteína [2, 3]. Por conseguinte, a identificação de alterações nas redes PPI associadas à forma expandida da ataxina-3 pode lançar alguma luz sobre os mecanismos patológico característicos da SCA3.

Uma vez que a rede proteica da ataxina-3 é constituída por mais de 400 proteínas, apenas algumas proteínas apresentam evidências experimentais que apoiam o seu papel na SCA3 [4, 5]. Além disso, esta rede de interações foi obtida utilizando metodologias como a espectrometria de massa e a Yeast Two Hybrid (Y2H), e é possível que nem todas as proteínas reportadas interajam com a ataxina-3, uma vez que estas metodologias são conhecidas por produzirem interações falso-positivas [6]. Por conseguinte, as metodologias *in silico* utilizadas neste trabalho podem ser utilizadas para validar a rede proteica da ataxina-3, bem como para identificar as proteínas que interagem com esta e que se comportam de forma diferente dependendo se é com formas selvagem (wt) e expandida (exp) da ataxina-3 que interagem. Para validar a rede de proteínas da ataxina-3, utilizei uma abordagem comparativa utilizando as diferentes bases de dados disponíveis no EvoPPI3 [6], bem como as regiões de interação obtidas utilizando a metodologia *in silico* empreendida por Rocha et al. [3]. Embora em Rocha et al. [3] as inferências da estrutura 3D se baseiem no I-TASSER, aqui também realizo um trabalho *in-silico* utilizando as estruturas 3D disponíveis na base de dados AlphaFold, para analisar como as inferências são afetadas pela utilização de um software de previsão diferente. Para identificar as possíveis proteínas que interagem com a ataxina-3 e que estão envolvidas na SCA3, identifiquei as proteínas que mostram interações fortes com a forma exp da ataxina-3 em comparação com a forma wt. Os resultados obtidos das duas previsões de estruturas 3D mostraram estar correlacionados e são identificadas as mesmas regiões de interação com a ataxina-3. No entanto, o número total de interações da ataxina-3 é maior quando se utilizam as previsões da estrutura AlphaFold. Além disso, a identificação de proteínas que se ligam de forma diferente com as formas wt e exp da ataxina-3 é afetada pela previsão da estrutura 3D utilizada e, por isso, para se ter confiança, apenas as proteínas que mostram resultados semelhantes com as duas metodologias devem ser consideradas

como as que importam na SCA3, devendo ser validadas experimentalmente. Neste trabalho também faço inferências biológicas para as ambas as conformações da ataxina-3 obtidas com as metodologias I-TASSER e AlphaFold. As inferências de acoplamento (docking) utilizando a previsão da estrutura AlphaFold da ataxina-3 produziram interações que diferem significativamente das descritas na literatura, mas que estão de acordo com as obtidas utilizando a previsão da estrutura I-TASSER. Estas conclusões são discutidas no contexto de trabalhos futuros.

Palavras-chave: ataxin-3, SCA3, expansão do trato polyQ, interações proteína-proteína, bases de dados de proteínas que interagem, previsões *in silico*

Abstract

Spinocerebellar ataxia type 3 (SCA3) is a neurodegenerative disease caused by the expansion of ataxin-3 polyglutamin (polyQ) tract [1]. This expansion of the trinucleotide repeat CAG leads to structural alteration of the encoded protein, that results on alterations in the native protein-protein interactions (PPIs) network associated with the regular activity of the protein [2, 3]. Therefore, the identification of PPI alterations associated with the expanded form of ataxin-3 may shed some light on the SCA3 pathological mechanism.

Because the ataxin-3 protein network comprises more than 400 proteins, experimental evidence supporting their role in SCA3 is available for only a few proteins [4, 5]. Moreover, the ataxin-3 network has been obtained using methodologies such as mass spectrometry and yeast two-hybrid (Y2H), and it is possible that not all ataxin-3 interactors interact with ataxin-3, since these methodologies are known to produce false positive interactions [6]. Therefore, *in silico* methodologies as used in this work, can be used to validate the ataxin-3 protein network as well as identify those ataxin-3 interactors that behave differently with the wild-type (wt) and expanded (exp) ataxin-3 forms. To validate the ataxin-3 protein network, I used a comparative approach using the different EvoPPI3 databases [6], as well as the interaction regions obtained using the *in-silico* methodology of Rocha et al. [3]. Although in Rocha et al. [3] the 3D structure inferences are based on I-TASSER, here I also perform the in-silico work using the 3D structures available in AlphaFold database, to address how inferences are affected by the use of a different prediction software. To identify putative ataxin-3 interactors involved in SCA3, I identify those ataxin-3 interactors that show strong interactions with the exp ataxin-3 form compared with the wt ataxin-3 form. The results using the two 3D structure predictions are correlated, and the same ataxin-3 interacting regions are identified. Nevertheless, the total number of ataxin-3 interactions is larger when using AlphaFold structure predictions. Moreover, the identification of proteins that bind differently with wt and exp ataxin-3 forms is affected by the 3D structure prediction used, and thus to be confident, only those proteins that show similar results with the two methodologies should be considered as those that matter in SCA3 and should be experimentally validated. In this work I also make useful biological inferences for the two ataxin-3 conformations obtained when using I-TASSER and AlphaFold methodologies. Docking inferences using the ataxin-3 AlphaFold structure prediction yielded interactions that significantly differ from those described in the literature, that are in agreement with those

obtained when using I-TASSER structure prediction. These findings are discussed in the context of future work.

Keywords: ataxin-3, SCA3, polyQ stretch expansion, protein-protein interactions, interactors databases, *in silico* predictions

Index

Agradecimentos.....	ii
Table List.....	x
Figure list.....	xi
Abbreviation list	xvi
Introduction.....	1
1. SCA3	1
2. ATXN3 and ataxin-3 protein	2
3. Pathogenesis	3
4. <i>In silico</i> approach	4
4.1 Ataxin-3 structure.....	4
4.1.1 Alpha Fold.....	5
4.1.1.1 I-TASSER.....	6
4.1.2 Predicting ataxin-3 Network.....	6
4.1.2.1 EvoPPI	7
4.1.2.2 PantherDB.....	8
4.1.2.3 Protein Docking	8
4.1.2.3.1 HADDOCK	10
Objectives.....	12
Chapter 1: Ataxin-3 Interactor database.....	13
1. Materials and Methods	14
1.1 The ataxin-3 interactome	14
1.2 Interactors functional characterization.....	14
1.3 Statistical analysis	15
2. Results and Discussion	15
2.2 Comparing human ataxin-3 interactors present in at least one interactome or predicted from other organisms' interactome database from those present in main databases only	21

2.3 Recognition regions are different in the ataxin-3 interactors are present in human interactors main databases and interactomes and/or predict interactomes from other model species and those that are in human interactors main databases only. 22

Chapter 2: Comparing ataxin-3 wild-type (wt) and expanded (exp) interactions using different interactors 3D structures predictions 24

1. Materials and Methods 25

1.1 Docking and interaction predictions. 25

1.2 Differences in 3D protein structures 26

1.3 Defining Binding regions 26

1.4 Statistical analysis 26

2. Results and discussion..... 26

2.1 Ataxin-3 interactions using I-TASSER and AlphaFold 3D structures predictions are highly correlated..... 26

2.2 Number of interactions in the JD and C-terminal regions are affected by the 3D protein structure used only in a small number of interactors 29

2.3 Identification of the ataxin-3 interactors affected by the presence of an expanded polyQ 30

2.4 Ataxin-3 interactors presenting six or more interactions in the polyQ region affect the results of the comparison of binding strength of ataxin-3 with exp and wt forms, when using I-TASSER and AlphaFold 3D structure predictions 31

Chapter 3: Using new interactors to evaluate their interacting pattern and predict ataxin-3 interaction regions 33

1. Material and Methods..... 34

1.1 Comparison between the interactors of the previous dataset and the new dataset 34

1.2 Detecting different behavior between proteins with different evidence for being ataxin-3 interactors..... 34

1.3 Tracing the proteins that bind strongly to the exp ataxin-3 than to the wt ataxin-3 form 34

2. Results and Discussion 35

2.1 Analyzing possible differences in the interaction patterns of ataxin-3 wt when using the new selected interactors 35

2.2	Detecting different behavior between proteins with different evidence for being ataxin-3 interactors.....	39
2.3	Detecting the proteins that are more prone to bind to the EXP ataxin-3 protein	41
Chapter 4: Using an AlphaFold ataxin-3 predicted structure for a docking simulation . 47		
1.	Materials and methods	48
1.1.	Addressing the similarities between both IT and AF predicted ataxin-3 structures with different the experimental ataxin-3 structures	48
1.2.	Docking of the interactors and evaluations of the interactions.....	48
2.	Results and Discussion	48
2.1.	I-TASSER and AlphaFold 3D structure predictions show similarity with different JD PDB structures.....	48
2.2.	The use of the Ataxin-3 AlphaFold structure suggests that the majority of the interactors interact more at the Ataxin-3 C-terminal region than at the JD, in contrast with experimental evidence	49
Conclusion and Future work		50
Bibliographic references		53
Appendix.....		60

Table List

Table 1. PHANTHER protein class enrichment analyses of the ataxin-3 interactors that are present in at least one interactome or predict interactome database (N= 237) with the EvoPPI3 ataxin-3 interactors only present in human main databases (N= 185)	21
Table 2. Interactors that show a higher number of interactions with the exp than the wt ataxin-3 forms using I-TASSER and AlphaFold 3D protein structures.....	30
Table 3. Proteins that mainly bind to the JD	35
Table 4. Comparison between the two datasets protein classes.....	39
Table 5. Comparison between the two datasets Molecular functions.....	41
Table 6. Proteins that show an increase in their interactions relative to the ataxin-3 wt when interacting with the exp form (in grey are those proteins that have been removed because they presented more than 6 interactions in the polyQ region; the values inside the () represent the number of ataxin-3 interactions excluding the polyQ region)	42
Table 7. Proteins that show an increase in ataxin-3 exp interactions higher than 10% when using the I-TASSER ataxin-3 predicted structure with I-TASSER interactors predicted structures (In grey are located the proteins whose results are in agreement to the ones presented in the IT-AF docking; the values inside the () represent the number of ataxin-3 interactions excluding the polyQ region)	43
Table 8. Proteins that presented an increase in ataxin-3 exp independently of the methodology used (ataxin-3 I-TASSER structure vs interactors I-TASSER structure; ataxin-3 I-TASSER structure vs interactors AlphaFold structures)	44
Table 9. TM scores for the proteins when considering the different 3D structure prediction methods (in grey are selected the proteins that presented an increase in ataxin-3 exp independently of the methodology used)	45
Table 10. TM score calculation between the ataxin-3 AlphaFold and I-TASSER predicted structures with ataxin-3 structures available at PDB database	49

Figure list

Figure 1 - Number of ataxin-3 interactors reported on EvoPPI (A) that show expression in basal ganglia (B), cerebral cortex (C), medulla oblongata (D), midbrain (E), pons (F), and thalamus (G), the 6 main tissues where SCA3 neurodegenerations is more severe	15
Figure 2 - Overlap of the <i>H. sapiens</i> ataxin-3 interactors on main databases (A) and the <i>H. sapiens</i> interactors reported in human polyQ_22 database	17
Figure 3 - Overlap of the 252 <i>H. sapiens</i> interactors on main databases also reported in <i>H. sapiens</i> polyQ_22 database (A) and the predicted interactome based on <i>D. melanogaster</i> modifiers database considering DIOPT (B) and Ensembl (C) orthologies	17
Figure 4 - Overlap of the selected 252 <i>H. sapiens</i> interactors on main databases also reported in human polyQ_22 database (A) and the predicted interactome based on <i>D. rerio</i> polyQ_22 database considering DIOPT (B) and Ensembl (C) orthologies5.....	17
Figure 5 - Overlap of the human ataxin-3 interactors present in the human polyQ_22 database and <i>D. melanogaster</i> (those identified using both DIOPT and Ensemble orthologies (A) and those identified using only DIOPT or Ensemble orthologies (B) and <i>D. rerio</i> interactors in both both DIOPT and Ensemble orthologies (C) DIOPT orthologies (D))	18
Figure 6 - Analysis of the ataxin-3 interactors reported by Sousa e Silva et al [106] and the proteins expressed in the 6 main tissues where SCA3 neurodegenerations is more severe A Reported <i>H. sapiens</i> interactors on the main databases B Basal ganglia expressed pro.....	18
Figure 7 - Analysis of the selected <i>H. sapiens</i> interactors and the <i>H. sapiens</i> interactors reported in HomoPolyQ databases A <i>H. sapiens</i> selected interactors B HomoPolyQ reported ataxin-3 interactors	18
Figure 8 - Overlap of the <i>H. sapiens</i> ataxin-3 interactors reported by Sousa e Silva, et al [106] that show expression in tissues relevant to SCA3 reported in human polyQ_22 database (A) and the predicted interactome based on <i>D. melanogaster</i> modifiers database considering DIOPT (B) and Ensembl (C) orthologies.....	19
Figure 9 - Overlap of the <i>H. sapiens</i> ataxin-3 interactors reported by Sousa e Silva, et al. [106] that show expression in tissues relevant to SCA3 reported in human polyQ_22 database (A) and the predicted interactome based on <i>D. rerio</i> predict interactomes databases considering DIOPT (B) and Ensembl (C) orthologies.....	20

Figure 10 - Overlap of the <i>H. sapiens</i> ataxin-3 interactors reported by Sousa e Silva, et al [106] that show expression in tissues relevant to SCA3 reported in human polyQ_22 database the and the predicted interactome based on <i>D. melanogaster</i> modifiers database considering both DIOPT and Ensembl (A) and DIOPT (B) orthologies, and predicted interactome based on <i>D. rerio</i> predict interactomes databases considering DIOPT and Ensembl (C) and DIOPT (D) orthologies.	20
Figure 11 . Percentage of interaction per protein residue at the five molecular recognition regions (IR) reported for proteins that interact mostly at the Josephin domain, according to Sousa e Silva et al. [106], for proteins that ataxin-3 interactors are present in human interactors main databases and interactomes and/or predict interactomes from other model species (upper panel, in black) and those that are in human interactors main databases only (lower panel, in grey), respectively.	23
Figure 12 . Pearson (R) linear correlation of the number of interactions along ataxin-3 when using 132 interactors, and the 3D protein structures prediction from AlphaFold and I-TASSER.	27
Figure 13 . Percentage of interaction per protein residue at the ataxin-3 for 132 interactors using AlphaFold predicted 3D structures (upper panel, in blue) and I-TASSER 3D structures (lower panel, in orange). Green dotted boxes represent regions oof interaction where more than 50 percent of the proteins show interaction with ataxin-3. The Josephin domain (JD) is assigned as a dotted box, while the NES77 and NES141 regions are marked with white boxes. UIM regions are also assigned within grey boxes	28
Figure 14 . Pearson (R) linear correlation of the number of interactions considering the interacting regions (those marked with dotted green boxes in Fig. 14) in common between AlphaFold and I-TASSER 3D protein structures predictions.	28
Figure 15 . Overlap of the ataxin-3 interactors that show a larger number of interactions in the JD when considering AlphaFold (A) and I-TASSER (B) 3D structure predictions and those interactors that show a large number of interactions in the C-terminal region when considering AlphaFold (C) and I-TASSER (D) 3D structure predictions, independently of considering the number of interactions in the polyQ region (see Supplementary Table 6 for Genelds).	29
Figure 16 . Number of interactions at the polyQ region using AlphaFold (upper panel) and I-TASSER (lower panel) for the ataxin-3 wt (in blue) and exp (in orange) form	32
Figure 17 . Comparison between the interaction precentages of each amino acid for ataxin-3 wt form, for both the new dataset (upper panel, blue color) and the previously	

studied dataset (lower pannel, orange color). In green dotted boxes are defined the interaction regions 37

Figure 18. Pearson (R) linear correlation of the percentage of interactions along ataxin-3 when using 3D protein structures prediction from AlphaFold for 83 ataxin-3 interactors here analysed and 100 ataxin-3 interactors analysed in Chapter 2, that have in common higher frequency binding with the JD region 38

Figure 19. Pearson (R) linear correlation of the percentage of interactions along ataxin-3 of the interactors with more evidence for being ataxin-3 interactors and less evidence for being ataxin-3 interactors.....38

Figure 20. Comparison between the interaction percentages of each amino acid for ataxin-3 wt form, for both the dataset of interactors with less evidence for being ataxin-interactors (upper panel, blue color) and the dataset of interactors with more evidence for being ataxin-3 interactors (lower pannel, orange color). In green dotted boxes are defined the interaction regions.....39

Supplementary Table list.

Supplementary Table 1. Representation of Ataxin-3 interactors in the main databases	60
Supplementary Table 2. Ataxin-3 human interactors present in different EvoPPI3 predict interactomes databases	63
Supplementary Table 3. Ataxin-3 human interactors present in main databases, excluding the 151 in Sousa e Silva et al. [106], that are also expressed in reference SCA3 tissues, present in human polyQ database (N=252) in different EvoPPI3 predict interactomes databases	77
Supplementary Table 4. The distribution of the 81 ataxin-3 human interactors analyzed in Sousa e Silva et al. [106], that are expressed in reference SCA3 tissues and present in human polyQ database in different EvoPPI3	78
Supplementary Table 5. The distribution of the 67 ataxin-3 human interactors analyzed in Sousa e Silva et al. [106], that are expressed in reference SCA3 tissues but not present in human polyQ database in different EvoPPI3 predict interactomes databases	79
Supplementary Table 6. Proteins that present the same interaction behavior when using ataxin-3 I-TASSER predicted structure (IT-IT) and AlphaFold predicted 3D structures of predicted ataxin-3 interactors	80
Supplementary Table 7. Interactors that mainly bind to the C-terminal region of wt ataxin-3 form using I-TASSER structures and AlphaFold interactors structure predictions	82
Supplementary Table 8. Interacting regions for each wt dataset and their correspondent interaction percentage (For the statistical analysis comparing the different interaction regions of each group, additional regions were added for both samples to have the same size. These additions do not reflect molecular signatures of the specific dataset and are marked in blue)	82
Supplementary Table 9. Proteins with different supporting evidence for being ataxin-3 interactors	85
Supplementary Table 10. Selection of the proteins that bind to the JD with different supporting evidence for being ataxin-3 interactors	85
Supplementary Table 11. Interacting regions and their correspondent interaction percentage for each group of proteins with more or less evidence for being ataxin-3 interactors (For the statistical analysis comparing the different interaction regions of each group, additional regions were added both samples to have the same size. These	

additions do not reflect molecular signatures of the specific dataset and are marked in blue) 86

Supplementary Table 13. Differences of the number of interactions along expanded and wild type ataxin-3 forms when using I-TASSER and AlphaFod 3D protein structures considering the total number of interaction sites including the polyQ region and excluding it, as well as the percentage of the number of interaction residues with expanded form compared with wt form and the total number of interactions with wt ataxin-3 88

Abbreviation list

AIR: Ambiguous Interaction Restraints
ATXN3: Ataxin-3
ATXN1: Ataxin-1
DRPLA: Dentatorubral-Pallidoluysian Atrophy
CASP: Critical Assessment of Protein Structure
CPORT: Consensus Prediction Of Interface Residues in Transient
exp: Expanded
GO: Gene Ontology
HADDOCK: High Ambiguity Driven protein-protein Docking
IT-IT: I-TASSER-I-TASSER
I-TASSER: Iterative Threading ASSEMBLY Refinement
IT-AF: I-TASSER-AlphaFold
JD: Josephin-Domain
MJD: Machado Joseph Disease
NCBI: National Center of Biotechnology Information
PANTHER: Protein Analysis Through Evolutionary Relationships
polyQ: polyglutamine
PPI: Protein-Protein Interactions
SCA3: Spinocerebellar Ataxia Type 3
SBMA: Spinal and Bulbar Muscular Atrophy
UIM: Ubiquitin Interacting Motif

Introduction

Estimations indicate that approximately 50-70% of the human genome is comprised of repetitive regions. In these regions, we can identify trinucleotide repeat sequences that, although common in the human genome, may present an expansion in their numbers leading to the emergence of diseases [7].

Since these regions are prolific in the human genome, they can be located in a large variety of regions, and their expansion can be located both in non-coding sequences or in exonic repeats (CAG) that code for polyglutamine tracts [7].

The later subclass is denominated polyglutamine (polyQ) diseases, which comprises 9 rare, inherited, and fatal neurodegenerative diseases such as Huntington's disease, six spinocerebellar ataxias, spinal and bulbar muscular atrophy (SBMA), and dentatorubral-pallidoluysian atrophy (DRPLA) [8]. All of them share some key features, such as being autosomal dominant diseases (with the exception of SBMA), they affect the central nervous system, display an accumulation of disease protein aggregates (mainly in neurons), have a slow progression over the years and finally, are typically fatal without an approved treatment to alter the course of the disease [1].

Among these diseases, Spinocerebellar Ataxia Type 3/Machado Joseph Disease (SCA3/MJD) is the most frequent ataxia worldwide [9] and is the main study topic of this project.

1. SCA3

As previously mentioned SCA3/MJD is a neurodegenerative disease characterized by its heterogenicity of clinical features such as its progressive Ataxia, ocular movement abnormalities and short life expectation. The disease mainly affects the brainstem, the spinal cord and the cerebellum leading to a prominent neurodegeneration in these areas[10].

Caused by an expansion of the trinucleotide repeat (CAG), located on the 10th exon of the *ATXN3* gene the affected individuals can present repeats ranging from 56-87, whereas an healthy patient the polyQ tract ranges from 12-44 CAG repeats [1].

With varying lengths in the expansion, the SCA3 disease leads to a phenomenon referred to as genetic anticipation, characterized by an inverse correlation between the age at onset and the size of the CAG repeats present in the pathogenic allele [11]. The

CAG repeat length is estimated to determine approximately 50-60% of the variability in age at onset [12], leading to a range of ages from 10-78 years, where the majority of the cases emerges at the age of around 40 years [13]. Furthermore, the life expectancy of a patient can also be highly variable, ranging from 7-29 years after the onset of the disease, with a mean survival time of 21 years [14]. The length of the expansion has additionally a major influence on the phenotype of the SCA3. According to Riess *et al* [15], repeats of more than 73 CAG motifs are often associated with a pyramidal phenotype presenting ataxia and spasticity (motor disorder where certain reflexes in our muscles become exaggerated when stretched quickly), in contrast with patients who present less than 73 CAG repeats and are more likely to develop neuropathy.

Another important hallmark of the SCA3, is the presence of intraneural inclusions, that contain the mutant ataxin-3 protein. This characteristic is common to other polyQ diseases, however, the full understanding of its role in the underlying pathogenic process of the disease is unknown[16].

2. ATXN3 and ataxin-3 protein

The gene *ATXN3* consists of 11 exons that encodes a protein called ataxin-3, with a molecular mass of 42-kDa that when mutated, turns out to be the main responsible for the SCA3 disease. Its structure, is composed of an N-Terminal Josephin domain (JD), two ubiquitin interacting motifs (UIMs), the poly Q stretch, and a variable C-terminal tail [16-18]. As for the Josephin Domain, it is constituted of approximately the first 180 residues, being the core of the enzyme's catalytic activity. This region as well as the C-Terminal region display conformational plasticity[18].

This protein is ubiquitously expressed and is known to be a deubiquitinating enzyme involved in the cellular protein quality control system. There have been identified different protein isoforms of the ataxin-3 that arise from the process of alternative splicing. This process originates mainly three different isoforms: ataxin-3aL, ataxin-3aS, and ataxin-3c, and as mentioned above, these isoforms mainly differ in their C-terminal sequence. This region on the first two isoforms emerges as a hydrophobic structure, while the ataxin-3c contains a hydrophilic region with a third UIM [19]. However, it has been reported that the presence of 2 or 3 UIM displays similar *in vitro* deubiquitinating activity [20]. Nonetheless, a higher variety of isoforms (56 alternative splicing variants) have been described by [21], some presenting alternative splicing in the N-terminal region, being differentiated by the Josephin Domain. Amongst all these different isoforms, ataxin-3c is the most predominant expressed in the brain [19, 20].

The protein is normally located in both cytoplasm and nucleus, however, in the disease mutated form, ataxin-3 [19] tend to be located in large concentration on the nucleus where it forms aggregates [22].

Ataxin-3 is mainly known for regulating E3 ligases, participating in the delivery of proteasomal substrate interacting with shuttle proteins, and regulating the aggresome formation. In addition, ataxin-3 was also shown to be capable of binding to DNA and interacting with transcription regulators, and in this way, have a role in transcriptional regulation [4, 5]. Based on this information, it is expected that ataxin-3 interacts with a wide variety of proteins responsible for diverse biochemical pathways, which represents a risk of dysregulation on several fronts in case the ataxin-3 suffers from a mutation, such as an expansion on the polyQ tract.

3. Pathogenesis

In spite, of the *ATXN3* been identified as a clear cause of the SCA3 disease, the pathological mechanisms underneath it are still unclear for researchers [22, 23]. Nonetheless, the underlying mechanism of the disease involves both changes in the localizations of the ataxin-3 expression and its function [24].

The hallmark of SCA3/MJD is the formation of protein aggregates that contain the pathological expanded ataxin-3. Though polyQ expansion is essential for the onset of the disease, the Josephin domain itself is strongly prone to aggregate, furthermore, the full-length ataxin-3 or its calpain and caspase cleavage fragments, show capability of aggregation, leading to suggest that the ataxin-3 protein contains multiple aggregation motifs [25], and both normal and expanded proteins are propense to form aggregates. However, the mutant enzyme, unlike the normal ataxin-3 has a higher susceptibility to form aggregates [26].

It is suggested that this propensity to form aggregates of the mutated protein, could arise from conformational changes induced by the expansion of the polyQ tract [27-29]. This aggregation process is a highly complex phenomena which contemplates interconnected mechanisms, such as protein conformational changes and formation of inter-molecular bonds which can involve single residues, as also the full molecule structure [30]. All these factors are results of alterations of the native PPIs, associated with the regular protein activity [2, 3], and could lead to subtle alterations in the substrate specificity or in protein-protein interactions (PPI) that are not fully understood in the spectrum of the pathological mechanisms underneath the disease [1].

Although suffering and expansion in the polyQ tract and forming aggregates containing mutated ataxin-3, it was reported by Wilborne, *et al* [31], that the mechanism requiring the ubiquitin-related properties of the ataxin-3 protein, are retained both in the ataxin-3 wild-type and in the expanded form, which is consistent with their observation that both protein forms retain the enzymatic activity. Furthermore, both structures were capable of binding to the poly-ubiquitin chain, also showing that the expansion does not prevent the binding of the expanded ataxin-3 to its substrate.

4. *In silico* approach

4.1 Ataxin-3 structure

As already mentioned, ataxin-3 is comprised of both the JD and its flexible C-Terminal region. However, the full structure of the protein at the moment is not fully determined, with exception of its JD, where three different structures are available [32].

To determine the full 3D structure of a protein, experimental techniques such as x-ray crystallography can be used, with the disadvantage of being a laborious, expensive and time consuming technique [33].

Nonetheless the fully amino acid sequence of the protein is available in databases such as the NCBI (National Center of Biotechnology Information) [34], allowing for the prediction of full 3D structures, using these sequences. Such predictions have been a challenge in computational biophysics in the last decades, leading to a panoply of prediction algorithms, from genome interpretation to protein function prediction [35]. Many advances in protein structure prediction and design are in paired with technological breakthroughs, at the same time as biological databases have been rapidly growing, however, such protein-modelling algorithms are computationally demanding [35], being many of these algorithms available in online servers.

There are two main approaches to the prediction of a protein structure: a template-base modelling and a template-free modeling [35]. The first approach, as the name suggests, uses a protein structure as a template for the prediction. For this, a representative protein structure database is used in order to search and identify a protein structurally similar to the target protein, therefore being its template. Next, an alignment between the target and its template is in order, where equivalent residues should be aligned together. Finally, based on this alignment and the template structure the predicted structure is built and the model quality evaluated [36].

For template-free modeling, the prediction comprises algorithms/pipelines/methods that generate models without any known structural homologs available. This approach mainly focus on physics based principles and energy terms, where the algorithms are designed with the ability to rapidly located global energy minimum and that are capable of selecting the best predicted structure based on a scoring function, among several generated models [37].

Based on these two approaches, a third one is possible to implement. An Hybrid approach that allows for an amalgamation of both template-based and template-free methods instead of using the pure criteria of one [37].

I-TASSER (Iterative Threading ASSEMBly Refinement) [38] and AlphaFold [39], are two of these protein structure prediction methods, that are available online, and were both used to predict structures from both the ataxin-3 structures (wild type (wt) and expanded (exp)) and its different interactors.

4.1.1 Alpha Fold

AlphaFold is a protein structure prediction approach developed by Jumper, *et al* [39] that is capable of providing a 3D protein structure with high accuracy even in cases where a protein model is not available. This success in accuracy was assessed in the 14th edition of the Critical Assessment of Protein Structure (CASP), where is determined the state of art in protein structure modelling.

In this community experiment, participants are provided with sequence of amino acids from 98.5% target proteins of the human proteome, and 3D models are built and compared with experimental structures obtained by independent assessors [40, 41]. Based on this analysis, AlphaFold was presented as vastly more accurate than the other competing approaches[39, 42]. Of the 98.5% of the human proteome covered, 58% of the residues heave been modelled with high confidence scores, presenting a positive scenario in protein structure research since less than 10% of the human proteome is covered by experimental results[43]

AlphaFold, utilizes deep learning [44], a methodological toolkit utilized in artificial intelligence, that allows to build neural networks in a multilayer way, that allows machines to learn from experience, without being endowed by the researcher [45]. This methodology is used to generate scoring functions from experimentally determined protein structures, by comparing its similarity to the native conformation [44, 46].

Furthermore, the usage of sequence alignments to predict a structure is another approach imbued in the architecture of AlphaFold [39].

The access for this tool in protein structure prediction, can be accessed in the European Bioinformatics institute (<https://alphafold.ebi.ac.uk/>). Each protein available in the database as a dedicated structure page, that shows the basic information, allowing the user to consult its 3D structure as well as the confidence level that can vary alongside the chain. Nonetheless many proteins can present lower confidence scores in certain zones of the chain, which can be correlated with backbone flexibility and intrinsic disorder of the protein [47].

4.1.1.1 I-TASSER

In this project is I-TASSER approach [38] was also used for protein modeling. Since its inception in 2006, the I-TASSER algorithm has undergone several alterations in its architecture, leading to improved performances over the years.

Initially designed for generating high-resolution models of small proteins using the TASSER approach [48, 49], I-TASSER utilizes threading to predict the 3D structure of the protein by matching sequences with similar folds without the need for an evolutionary relationship, as opposed to comparative modeling, where an evolutionarily related and solved template is aligned with the target protein [49].

Over the years, the I-TASSER algorithm has undergone various updates and improvements, culminating in 2022 with an automated pipeline for multi-domain protein structure assembly and function named I-TASSER-MTD. This pipeline was built on the I-TASSER architecture incorporating deep learning algorithms, improving its ability and accuracy for domain modeling and inter-domain assembly [50].

I-TASSER server is available since 2006 on the I-TASSER server for protein structure and function prediction (<https://zhanggroup.org/I-TASSER/>), allowing non-experts to generate protein structure models for their own sequences deposited on the server. Using only the I-TASSER approach, the server generated the best predictions experiment outperforming all other automated servers on the community-wide blind experiment CASP7 [51].

4.1.2 Predicting ataxin-3 Network

As mentioned above, one of the pathogenic hallmarks of the SCA-3 disease is its aggregation both with itself, in normal or expanded form, and with other proteins. This

protein is involved in a variety of cellular pathways having a crucial role in ubiquitin proteasome system. Furthermore the protein is known for regulating expression various genes being also able to interact with other ataxin-3 molecules [19].

This presence in various pathways leads the protein to be involved in various PPIs, and many of its interactors have been obtained using high-throughput screen techniques such as mass spectrometry, two-hybrid assay and tandem affinity purification. These three techniques are the three most used experimental measurements techniques, and although its popularity, they present very high rates of false positives interactions, being able to reach rates of 50 % [6]. Such drawbacks added to the labor and costs of these experimental methods, are leading for an increased attentions and research in computational approach for PPIs prediction[52].

4.1.2.1 EvoPPI

These PPIs datasets are available for a large variety of species, and were determined using multiple detection approaches, such as the ones mentioned above. The access to these datasets is made by databases such as BioGRID[53], MINT[54] and HIPPIE [55] allowing these results to be downloaded. However, some problems arise from having so much data scattered by a large variety of databases, for example, although protein interactions are reported the comparison between databases is difficult since many of these present the information in different formats (gene identifiers (BioGRID), gene names (CCSB), among many others) [6, 56]. For these kind of problems, EvoPPI a web service available on <http://evoppi.i3s.up.pt> was developed in order to allow a comparison between these databases [6], since all the of the interactions declared on the database are reported as GeneID pairs.

Other problem arising from all data available, is that often PPI networks are very large, and although databases provide a large variety of information, this probably only covers a small fraction of the full network. From this EvoPPI, tries to break this barrier under the assumption that PPI networks are conserved between distantly related species[56, 57]. This allows to predict some interactomes based on interactomes presented by model species, making the most out of the available data. This web applications even features cross-species comparisons using a BLAST-based approach for these comaprison[58].

Additionally, along the years, the library of database haven been updated, leading to an extension in the number of unique interactions [59]. Therefore, EvoPPI can be used to track and select ataxin-3 interactors for the docking simulations performed in this work.

4.1.2.2 PantherDB

PANTHER (Protein Analysis Through Evolutionary Relationships) is a resource that allows the user to analyze the classification of certain genes according to their evolutionary history and their functions [60]. For the functional classifications, PANTHER makes use of Gene Ontology (GO), including both annotations provided by the Gene Ontology Consortium (available at: <http://geneontology.org>) and inferred annotations. Among the various analyses offered by PANTHER, “overrepresentation test” is one of the most useful for this work, allowing the users to identify enriched genes from their input list. The test compares the data with a reference gene list specific to the selected organism and performs a statistical analysis to determine over or underrepresentation of a gene in a particular class [60].

In order to conduct accurate analyses, the overrepresentation test from PANTHER has undergone several updates in its procedure. One significant change was the switch from a binomial test to a Fisher’s test [61] which allowed for more accurate analysis of smaller gene lists. Additionally, the statistical adjustment method was revised. Previously, Bonferroni correction [62] was used, however, the correction method was too conservative masking biologically significant results [60]. As a result, it has been replaced with False Discovery Rate correction [63].

4.1.2.3 Protein Docking

PPIs can range from a large variety of biological processes, including cell-to-cell interactions and metabolic and developmental control [64]. Proteins typically do not act as isolated particles when performing their functions on a living organism, so, studying these PPIs is a vital step for understanding the protein function within a cell [65]. Since many diseases as SCA3 involves genetic mutations that alter the functionality of a protein, by studying the PPI network, for scenarios like this, is possible to pinpoint functional consequences of genetic variations[66].

Although ataxin-3 has been intensively studied by its role in SCA3 disease, the full PPI network of this protein seems to be incomplete, with novel interactions being reported every year [67-69]. Having in consideration this lack of full disclosure on the ataxin-3 network as well as the limitations on experimental techniques, protein docking can be an alternative for the full exposé of the network.

Protein docking has its objective in predicting the near-native orientations of proteins, where one of them is designated as the receptor and the second one the ligand. These predictions have a role in delineating the characteristics of interactions, as well as the

role of a mutation in interaction patterns and affinity predictions. The process of docking generally involves two steps: the pose generation and the scoring of the structure. Nonetheless, different algorithms can lead to a high variety of procedures and consequently results [70]. In this first stage, regions of high shape and biochemical complementarity are searched in a 6-D space (three dimensions of translational freedom and 3 dimensions of rotational freedom making the protein move around this space. This step has been a critical component of the docking research since it is the development of novel techniques that can increase the speed of this search. For this stage many methods can be used by different softwares, but one of the most popular is the Fast Fourier Transform (FFT) [71]. As for the second stage, it involves assigning a score to the different proteins poses. The total score of a prediction may variate according to the scoring function selected and can have in consideration factors like electrostatic and van der Waals potential, hydrogen bonds, hydrophobicity, among many other criteria. These factors contribute for a higher or lower final score, and among the best classified structures, they can be selected or submitted to a refinement process [70].

There are two main types of docking: rigid body and flexible docking. The first method of docking as the name suggests treats the proteins used in the docking as solid objects, exploring a high variety of docking conformations evaluating the geometric fit or complementarity of surface for each individual structure [72]. One of the algorithms encompassed in the rigid-body methodologies is the FTT [71], where one protein is placed in the origin of a coordinate system on a fixed grid, while the second protein is placed on a movable grid. As for the interaction energy it is written as a few correlation function, and by combining with the ability of this algorithm to sample a significant number of conformations allows obtaining energy values at each grid point, not needing any prior information on the structure of the complex [73]. Nonetheless this kind of approach has its limitations. When docking co-crystallized protein conformations, the rigid-body methodology may be successful, however the scenario is reversed when using independently crystallized proteins, leading to wrong conformations [72, 73]. Other limitation is related with the fact that these methods rely on scoring functions that allow for overlapping results, limiting the accuracy of the docking [74].

As it was introduced by Daniel Koshland, proteins upon interacting can change its conformation, presenting a certain amount of plasticity of the active site that receives the ligand [75]. Such property present a barrier in predicting protein-protein complexes [76] since it requires a higher number of degrees of freedom, increasing both the running time as the number of false-positives[77].

For such scenarios, introducing flexibility in the docking can be used to facilitate the docking process, if there is previous knowledge of the flexible regions and conformational changes. This kind of information can be obtained by various methods like *in silico* simulations from molecular dynamics, however other approach could involve using several structures of protein or its homologues where information about its flexible and hinge regions have been solved [76].

Flexibility in the docking method can be introduced on different steps, either by smoothing the protein surfaces or allowing interpenetration in some degree or by using multiple dockings from various conformations. Lastly, there is the possibility of allowing sidechain and backbone flexibility both in the docking and refinement step [76].

4.1.2.3.1 HADDOCK

Throughout this work, HADDOCK (High Ambiguity Driven protein-protein Docking) server [78, 79], which can be accessed online at the BovinLab website (<https://bianca.science.uu.nl/haddock2.4/>) was used extensively as our docking method. HADDOCK employs biochemical and/or biophysical interaction data to produce the resulting protein dockings using the chemical shift perturbation data from NMR titration experiments or mutagenesis data. The presented information is in the form of Ambiguous Interaction Restraints (AIRs). It concerns the residues involved in protein interactions, separated into two categories: Active and Passive residues. The active residues correspond to all the residues showing a significant chemical shift perturbation upon the formation of the complex combined with a high solvent accessibility in the free-form protein. For mutagenesis data, the active residues are the ones that show up on mutations able to alleviate the complex formation and are also solvent exposed. As for the passive residues, they correspond to the residues that show a less significant chemical shift perturbation, being also surface neighbors of the active residues having a high solvent accessibility, nonetheless. The remaining residues that do not satisfy the high solvent accessibility are removed from the list [80].

The docking protocol comprises three distinct stages: (1st) randomized orientations and rigid body energy minimization, (2nd) semirigid/semi-flexible simulated annealing in torsion angle space, and (3rd) final refinement with explicit solvent. In the semi-flexible simulated annealing, the amino acids defined as active and passive residues used in the AIRs are allowed to move and optimize their interface packing. This allows for the sampling of a larger conformational space introducing flexibility to the docking protocol [80, 81].

As demonstrated, the strength of the HADDOCK approach lies in its ability to leverage experimental data to generate accurate protein dockings. However, one of the limitations of this approach has been the reliance on prior knowledge of the binding modes of the molecules involved. To address this challenge, CPORT (Consensus Prediction Of Interface Residues in Transient complexes), which utilizes various computational algorithms and consensus-based approaches was used to identify residues that are likely to participate in the protein-protein interface [82].

The aforementioned study demonstrated that the combined use of HADDOCK-CPORT achieved promising results. In the cases where the residue data had been experimentally determined, CPORT showed high accuracy, successfully predicting a portion of the active and passive residues. This breakthrough allowed HADDOCK to expand its applicability beyond the limitations of data-driven docking and approach the capabilities of *ab initio* docking methods combining the strengths of data-driven docking and interface prediction, demonstrating the potential to overcome previously traced limitations. By integrating CPORT with HADDOCK, a significant advancement was made in closing the gap between data-driven docking and *ab initio* docking methods.

Overall, HADDOCK demonstrated to be reliable docking approach. In CAPRI rounds 38-45, showed consistent high ranking indicating its effectiveness in predicting protein-protein interaction. The usage of data implemented in this approach improves the accuracy of the protein docking [83], nonetheless it's important to have in mind that the quality and reliability of this experimental data used as input can affect the accuracy of the docking.

Objectives

For the realization of this thesis, I've aimed for the following objectives:

1. Determine which of the proteins that have been reported as ataxin-3 interactors, may present an important role in the SCA3 pathogenesis;
2. Validate if the change of the 3D structure prediction method used for the interactors may affect the predictions made with the docking simulation;
3. Determine if the existence of 6 or more interactions in the polyQ region may or may not lead to misleading predictions;
4. Validate if the change of the ataxin-3 predicted 3D structure may affect the predictions made with the docking simulation.

Chapter 1: Ataxin-3 Interactor database

Studying Protein-protein interactions (PPI) has been made possible by utilizing high-throughput techniques such as yeast two-hybrid (Y2H) system, protein-fragment complementation assays, and affinity purification combined with mass spectrometry [52]. The data collected from such studies is compiled in primary databases such as BioGRID [84, 85], CCSB [86], Droid [87], FlyBase [88], HIPPIE [55], HitPredict [89], HomoMINT [90], INstruct [91], Interactome3D [92], Mentha [93], MINT [54, 94], and PINA v2 [95]), being all of them included in a meta-database, for PPI denominated EvoPPI3.

Since many of these databases present different formats, such as gene identifiers (BioGRID), UniprotKB accession numbers (MINT), and gene names (CCSB), comparing results from these databases can be arduous. However, by using EvoPPI, comparing these primary databases is easily performed by utilizing GeneID as identifiers. Despite being rigorously curated, false positive interactions may still persist in these databases due to the experimental limitations of the methods previously described. In addition, due to factors such as protein tissue and/or subcellular locations being responsible for genes to be expressed at different moments in time, unmatched physiological conditions, and post-translational modifications, this *in vitro* protein interactions may never be observed *in vivo*.

Given these factors, the inclusion of a specific PPI on a database may follow a varied range of criteria, leading to a limited data overlap between different databases where each of them can report an exclusive set of information [58]. Since the identification of false positive interactions has yet to come to a clear consensus, EvoPPI, addresses this issue by allowing its users to selective choose the databases used for searching and compare the reported PPIs enabling the user to decide which ones to trust.

Under the premise that the PPI networks are partially conserved between species (supported by the high conservation of human disease in the human and mouse genome (99.5%; [96]), and the presence of a significant number of paralogs for human disease-related loci in *D. melanogaster* and *Caenorhabditis elegans* (75% and 83% respectively) [97, 98]) and assuming that the homologous with similar functions reliably identified, it is possible to make comparisons between the networks obtained in different species. This approach allows the user to predict and test interactions not yet reported in other species, using the EvoPPI3 Predict-interactome databases. Additionally, for gene

conversion between different species datasets, selection can be made using DIOPT or Ensembl converters. Therefore, in this chapter, I will use interaction reports in other species to validate those available in EvoPPI3 for ataxin-3. In main databases, proteomic analyses from patients, human and mammalian derived cells, *D. melanogaster*, and *C. elegans* mutants are not included since novel protein interactions could be obtained when compared to the wild-type form. This is not the case for neurodegenerative polyglutamine (polyQ) where pathogenesis is primarily attributed to changes in protein binding strength that leads to the dysregulation of the protein network and not due to novel interactions [58, 99-102]. Therefore, for the nine polyq diseases, such data is available in EvoPPI3, in databases called PolyQ_22. Furthermore, large-scale genetic screening analyses using disease model species that recapitulate key features of the human polyQ diseases from vertebrate models such as *Mus musculus*, *D. rerio*, and non-vertebrate models such as *D. melanogaster*, and *Caenorhabditis elegans* are available as Modifiers_models_22. The two types of databases are also used to validate the interactions reported for ataxin-3.

1. Materials and Methods

1.1 The ataxin-3 interactome

The human ataxin-3 PPI were obtained using EvoPPI3 (<http://evoppi.i3s.up.pt>) [6] main databases by selecting the ataxin-3 (Gene ID: 4287), and level 1 interaction proteins (proteins that directly interact with the target protein). To validate the data from the main databases different datasets were collected using the three Interactomes main subcategories: Datasets, HomopolyQ (PolyQ_22) and Homomod (Modifiers_polyQ_22) [56]. Venn diagrams (<https://bioinformatics.psb.ugent.be/webtools/Venn/>) were used to identify common interactions on the different sets of data.

To address the expression on tissues of interest for SCA3 (Seidel, et al, [103]) data from Human Protein Atlas (<https://www.proteinatlas.org>) [104] proteins from the tissues where obtained.

1.2 Interactors functional characterization

The list of proteins with evidence of being ataxin-3 interactors, PANTHER [60] database was used for analysis of enriched molecular function using Gene Ontology analysis (available at: <https://www.pantherdb.org>).

Figure 1 - Number of ataxin-3 interactors reported on EvoPPI (A) that show expression in basal ganglia (B), cerebral cortex (C), medulla oblongata (D), midbrain (E), pons (F), and thalamus (G), the 6 main tissues where SCA3 neurodegenerations is more severe

identified with the GeneID 8022 has been reported in one main database, and all the others are reported in three databases (**Supplementary Table 1**).

Thus, the number of main databases reporting the ataxin-3 interaction seems not a good metric for interaction validation. Therefore, a comparative analysis was performed using polyQ interactomes and predicted interactomes to validate ataxin-3 interactors. Using human Interactome selection polyQ_22 (proteomic analyses of neurodegenerative polyQ patients or human-derived cell lines) for ataxin-3, 516 ataxin-3 interactors were obtained (**Supplementary Table 2**).

Additionally, using human Interactome selection modifiers_22 (from large-scale genetic screening analyses from human-derived cells and patients), other 15 ataxin-3 interactors were obtained. By performing the same analyses but using predicted interactomes from *Mus musculus*, no interactions in the main databases as well as in polyQ _22 (the predicted interactomes based on proteomic analyses of transgenic *M. musculus* expressing human ataxin-3) databases were found. For *M. musculus* modifiers database (based on large-scale genetic screenings using transgenic *M. musculus* expressing human ataxin-3), 1339 interactors were found using DIOPT and 547 interactors using Ensembl gene orthologies (**Supplementary Table 2**). When using data from *Danio rerio* predicted interactomes, no interactions were described in the main databases. In contrast, 1771 interactors were found in polyQ _22 databases using DIOPT, and 678 interactors were found using Ensembl gene orthologies. Performing the same analyses with *D. melanogaster*, 4 interactors were obtained when considering interactome selection polyQ_22 using DIOPT and 1 interactome considering interactome selection polyQ_22 using Ensembl gene orthologies (**Supplementary Table 2**). The same analyses using the Modifiers_22 database revealed 1911 and 1011 interactors considering DIOPT and Ensembl orthologies. When considering predicted interactomes from *C. elegans*, 15 ataxin-3 interactors were identified in the Modifiers_22 database when considering DIOPT orthologies and 7 ataxin-3 interactors when considering Ensembl orthologies (**Supplementary Table 1**). Considering these datasets of the 8 proteins that do not show expression in tissues that matter in SCA3, three (GeneID 3849, GeneID 347688, and GeneID 3305) have been reported as orthologous/paralogous of ataxin-3 interactors in Interactome PolyQ_22 from *D. rerio*, *M. musculus* and/or *D. melanogaster*, and thus were considered for further analyses. Only the five proteins not described as ataxin-3 interactors in any of the databases available at EvoPPI3 (GeneID 6407, GeneID 83889, GeneID 5430, GeneID 1828, GeneID 4085) were excluded from this study.

Of the 267 human ataxin-3 interactors, 252 have been reported in Interactomes polyQ_22 from *H. sapiens* (proteomic analyses of neurodegenerative polyQ patients or human and mammalian derived cell lines; **Figure 2**).

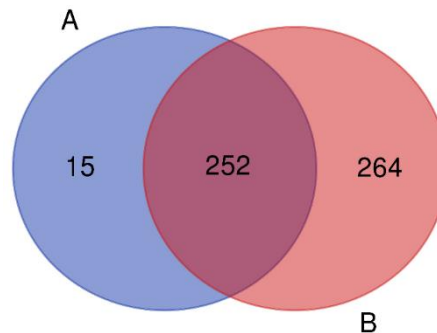


Figure 2 - Overlap of the *H. sapiens* ataxin-3 interactors on main databases (A) and the *H. sapiens* interactors reported in human polyQ_22 database

Of these 252 human ataxin-3 interactors and *D. melanogaster* predicted interactomes datasets, 48 have been described as *D. melanogaster* modifiers genes, using both DIOPT and Ensembl for the establishment of gene orthologies, 21 as *D. melanogaster* modifiers genes using DIOPT and 2 as *D. melanogaster* modifiers genes using Ensembl (**Figure 3**). Performing the same comparison with data from *D. rerio* predict interactomes databases (**Figure 4**), 117 human ataxin-3 interactors are present in PolyQ_22 database when considering DIOPT and Ensembl orthologies, plus 31 when considering DIOPT orthologies (**Figure 4; Supplementary Table 3**). The crossing of these datasets can be observed in **Figure 5**, which shows common interactors between the datasets of two predicted interactomes of two different species, which also cements the possibility of being ataxin-3 interactors.

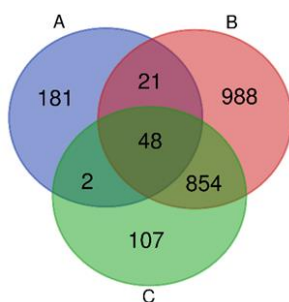


Figure 3 - Overlap of the 252 *H. sapiens* interactors on main databases also reported in *H. sapiens* polyQ_22 database (A) and the predicted interactome based on *D. melanogaster* modifiers database considering DIOPT (B) and Ensembl (C) orthologies

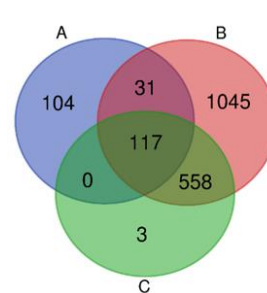


Figure 4 - Overlap of the selected 252 *H. sapiens* interactors on main databases also reported in human polyQ_22 database (A) and the predicted interactome based on *D. rerio* polyQ_22 database considering DIOPT (B) and Ensembl (C) orthologies

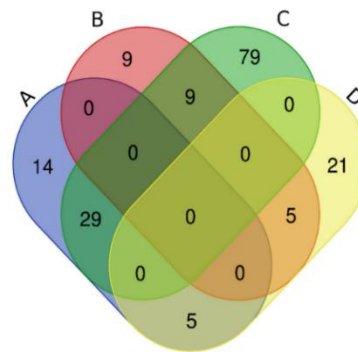


Figure 5 - Overlap of the human ataxin-3 interactors present in the human polyQ_22 database and *D. melanogaster* (those identified using both DIOPT and Ensemble orthologies (A) and those identified using only DIOPT or Ensemble orthologies (B) and *D. rerio* interactors in both both DIOPT and Ensemble orthologies (C) DIOPT orthologies (D))

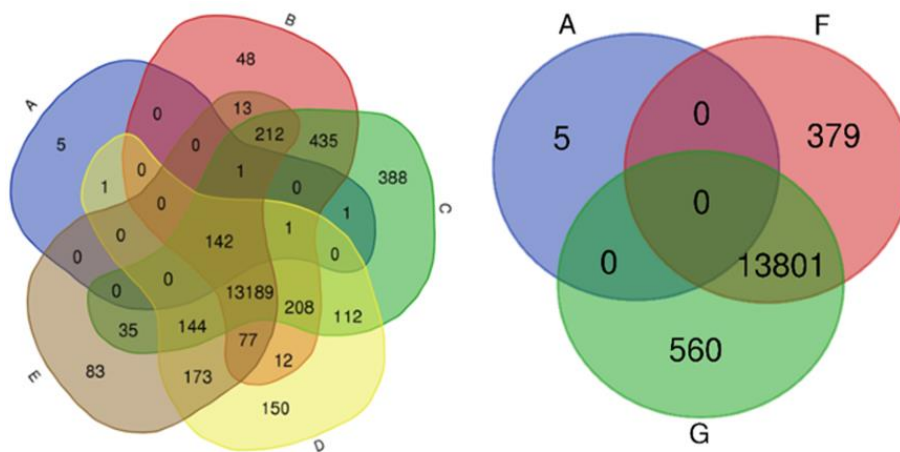


Figure 6 - Analysis of the ataxin-3 interactors reported by Sousa e Silva et al [106] and the proteins expressed in the 6 main tissues where SCA3 neurodegeneration is more severe A Reported H. sapiens interactors on the main databases B Basal ganglia expressed proteins

Applying the same methodology to the 151 proteins analysed in Sousa e Silva [106], all proteins except five (GeneID 56159, 57472, 84675, 387990, and 84676) are expressed in tissues where the SCA3 neurodegeneration is more severe (**Figure 6**). Of these five proteins, protein 56159 is present in H. sapiens polyQ_22 database and 387990 is

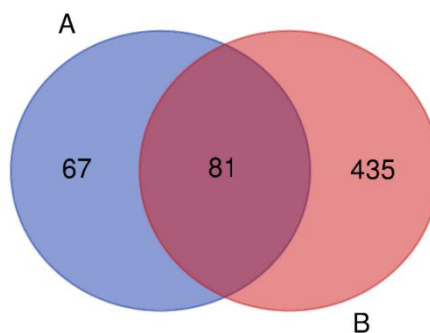


Figure 7 - Analysis of the selected H. sapiens interactors and the H. sapiens interactors reported in HomoPolyQ databases A H. sapiens selected interactors B HomoPolyQ reported ataxin-3 interactors

present in the *D. melanogaster* predict interactome using database of modifiers genes when using DIOPT and Ensembl orthologies (**Supplementary Table x2**). Therefore, the three proteins (57472, 84675, and 84676) not expressed in the tissues relevant for SCA3 and are not present in human polyQ databases could represent false ataxin-3 interactors. For the remaining 148 proteins, 81 are reported in human polyQ_22 database (**Figure 7; Supplementary Table 4**).

The 81 ataxin-3 interactors analysed in Sousa e Silva [106] reported in *H. sapiens* polyQ_22 database, 15 have been described as *D. melanogaster* modifiers genes using both DIOPT and Ensembl orthologies, and nine as *D. melanogaster* modifiers genes using DIOPT (**Figure 8; Supplementary Table 4**).

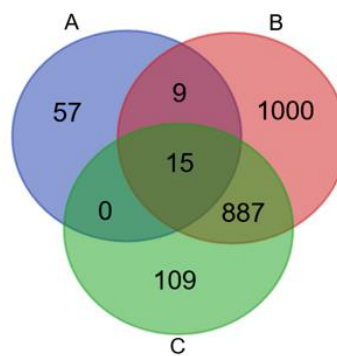


Figure 8 - Overlap of the *H. sapiens* ataxin-3 interactors reported by Sousa e Silva, et al [106] that show expression in tissues relevant to SCA3 reported in human polyQ_22 database (A) and the predicted interactome based on *D. melanogaster* modifiers database considering DIOPT (B) and Ensembl (C) orthologies.

Performing the same comparison with data from *D. rerio* predict interactomes databases, 22 human ataxin-3 interactors are present in polyQ_22 database when considering DIOPT and Ensembl orthologies, plus eight when considering DIOPT orthologies (**Figure 9; Supplementary Table 4**). The overlap of these four datasets of interactors is presented in **Figure 10 (Supplementary Table 4)**.

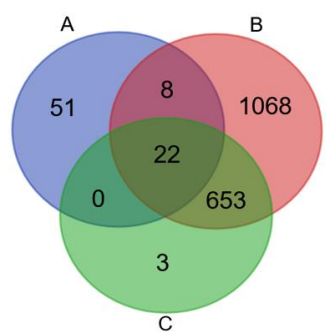


Figure 9 - Overlap of the *H. sapiens* ataxin-3 interactors reported by Sousa e Silva, et al. [106] that show expression in tissues relevant to SCA3 reported in human polyQ_22 database (A) and the predicted interactome based on *D. rerio* predict interactomes databases considering DIOPT (B) and Ensembl (C) orthologies.

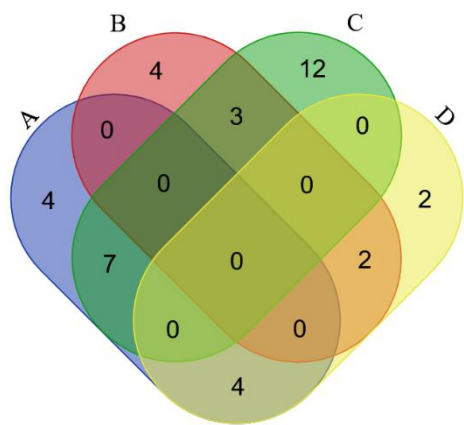


Figure 10 - Overlap of the *H. sapiens* ataxin-3 interactors reported by Sousa e Silva, et al [106] that show expression in tissues relevant to SCA3 reported in human polyQ_22 database the and the predicted interactome based on *D. melanogaster* modifiers database considering both DIOPT and Ensembl (A) and DIOPT (B) orthologies, and predicted interactome based on *D. rerio* predict interactomes databases considering DIOPT and Ensembl (C) and DIOPT (D) orthologies.

From the dataset used in Sousa e Silva et al [106], 16 human ataxin-3 interactors have been described in human polyQ_22, D.melanogaster, and D.rerio databases, and thus are here considered true interactors (**Supplementary Table 4**). I also considered as true interactor the 17 human proteins described in main databases, present in human polyQ_22 database that are also reported as interactor in D. rerio polyQ_22 database that are identified using DIOPT and/or Ensembl orthologies, as well as the eight proteins that have been identified as human proteins described in main databases, present in human polyQ_22 database that are also reported as interactor in D. melanogaster modifier polyQ_22 database using DIOPTan orthologies.

For the remaining 67 proteins analysed in Sousa e Silva et al. [106] that were not reported in the human polyQ_22 database, seven have been reported as D. melanogaster modifiers genes using both DIOPT and Ensembl orthologies, and five as D. melanogaster modifiers genes using DIOPT (**Supplementary Table 5**). When considering D. rerio predict interactomes databases, four are present in polyQ_22 database when considering DIOPT and Ensembl orthologies, eight when considering DIOPT orthologies, and three when considering Ensembl orthologies. Thus, these 22 unique proteins here considered as true ataxin-3 interactors (**Supplementary Table 4**). In summary, of the 151 proteins analysed in Sousa e Silva [106], using a comparative approach, I found evidence for 63 proteins being true ataxin3 interactors (see **Supplementary Table 4** and **Supplementary Table 5** for GeneIDs)

2.2 Comparing human ataxin-3 interactors present in at least one interactome or predicted from other organisms' interactome database from those present in main databases only

According to PANTHER protein class [60], all of the ataxin-3 interactors that are present in human main databases and in at least one other human interactome and/or predicted interactome database (N= 237) are significantly enriched ($P<0.05$) in proteases when compared with the EvoPPI3 ataxin-3 interactors only present in human main databases (N= 185), but the latter is enriched ($P<0.05$) in ubiquitin-protein ligases (**Table 1**).

Table 1. PHANTHER protein class enrichment analyses of the ataxin-3 interactors that are present in at least one interactome or predict interactome database (N= 237) with the EvoPPI3 ataxin-3 interactors only present in human main databases (N= 185)

Protein class	N=237	N=185	Fold Enrichment	P value
protease	26	4	5.24	2.24E-04
ubiquitin-protein ligase	4	23	.14	1.07E-05

The enrichment in proteases is in agreement with literature since a critical role of proteolysis has been described associated with the aggregation process observed in SCA3, although the identity of the protease(s) responsible for processing of AT3 is unknown [107]. Therefore, this dataset is likely enriched in true ataxin-3 interactors. Nevertheless, there is an excess of ubiquitin-protein ligases in ataxin-3 interactors only present in human main databases (**Table 1**) suggesting that not all of the 185 ataxin-3 interactors are false positives, and thus that the interactomes from other species may be incomplete. Ataxin-3 has a role in the ubiquitin-proteasomal pathway by rescuing proteins from degradation or by stimulating the breakdown by the removal of inhibitory poly-ubiquitin chains and by the regeneration of free and reusable ubiquitin.

2.3 Recognition regions are different in the ataxin-3 interactors are present in human interactors main databases and interactomes and/or predict interactomes from other model species and those that are in human interactors main databases only

Using a set of 132 wt ataxin-3 interactors, Sousa e Silva et al. [106] reported that 117 proteins that show more interaction residues (70.7%) at JD than at the remaining part of ataxin-3. These authors used the clusters of interaction sites to define molecular recognition patterns that can be used to identify novel ataxin-3 interactors *in-silico*, as well as the interaction type (mostly at the JD or outside this region). For those interactors that show most of the interaction sites at the JD they identify five molecular recognition regions at the JD, another at the end of the JD and adjacent region, and one at the end of the N-terminal region. For the set of 15 interactors that show more interaction sites at the C-terminal region of ataxin-3, six molecular recognition regions are identified, namely, one at the JD region, and five at the C-terminal ataxin-3 region. Since the proteins used can be divided in two groups, those that are also present in interactomes and/or predict interactomes from other model species, here considered as positive ataxin-3 interactors, and those that are present in human interactors main databases only, that we consider possible false positives, here I analyse if the two group of proteins behave differently according to the molecular recognition regions reported in Sousa e Silva et al [106] for those interactors that show most of the interaction sites at the JD (the largest data set, **Figure 11**). Indeed, the percentage of interacting sites in the molecular recognition regions are different (Sign test; $P < 0.05$). Nevertheless, the excess in the

percentage of interactions is observed in the group of interactors reported in human main databases only, those we consider possible false positives, suggesting that in this group there are true ataxin-3 interactors.

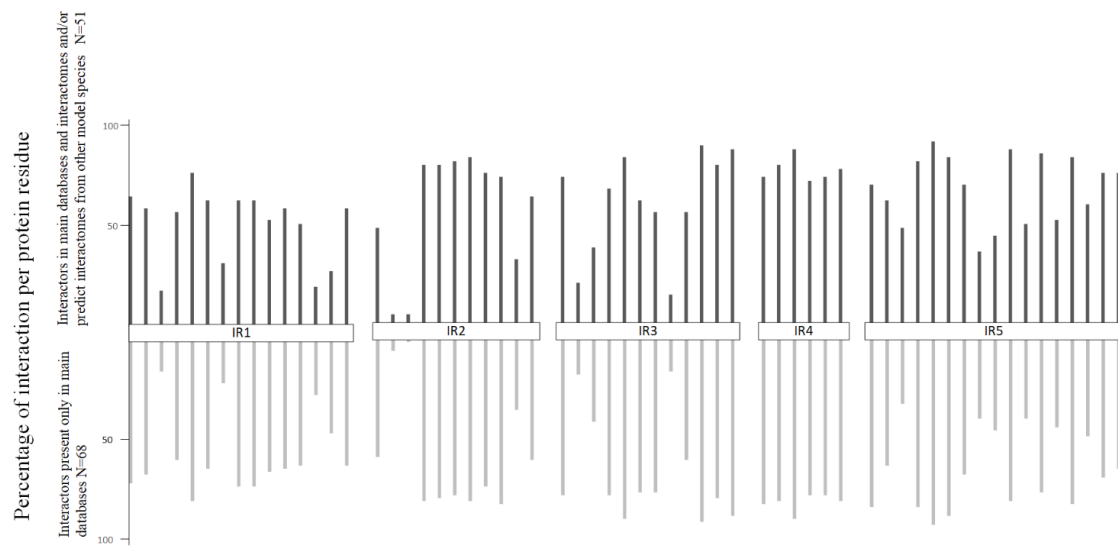


Figure 11. Percentage of interaction per protein residue at the five molecular recognition regions (IR) reported for proteins that interact mostly at the Josephin domain, according to Sousa e Silva et al. [106], for proteins that ataxin-3 interactors are present in human interactors main databases and interactomes and/or predict interactomes from other model species (upper panel, in black) and those that are in human interactors main databases only (lower panel, in grey), respectively.

In conclusion, the dataset is the ataxin-3 interactors are present in human interactors main databases and interactomes and/or predict interactomes from other model species are enriched in true ataxin-3 interactors but those interactors that are in human interactors main databases may contain also true ataxin-3 interactors.

Chapter 2: Comparing ataxin-3 wild-type (wt) and expanded (exp) interactions using different interactors 3D structures predictions

As previously mentioned, SCA3 pathology emerges as the result of a pathogenic expansion of the polyQ tract at the ataxin-3 protein. Although ataxin-3 has been identified as responsible for the disease, its pathological mechanisms are still unclear [22, 23]. The polyQ expansion can alter the protein's conformational state, affecting the interaction strength with its interactors, leading to alterations in the PPI [2, 3]. These alterations can be related with the aggregation process characteristic of the disease, since this process involves protein conformational changes and the formation of inter-molecular bonds that involve both single residues and the whole molecule structure [30]. The PPIs affected by the polyQ expansion are mostly unknown. Recently Sousa e Silva et al. [106] used an in-silico approach to address this question using a set of 151 ataxin-3 interactors. By comparing the strength of interaction of 132 (those for which the methodology worked) proteins with the wt (14Q) and exp (50Q) human ataxin-3 forms, they have identified 17 proteins that showed more than 10% increase in the number of interaction residues in exp ataxin-3 relative to the wt form. Sousa e Silva et al. [106] used the inferred I-TASSER [38] 3D structures for the ataxin-3 wt and exp forms and the interactors studied. These authors argue that although AlphaFold is considered the best tool for 3D structure prediction (CASP14 edition; [108] for ataxin-3 the only structure available is that of the wt ataxin-3, and thus, for consistency they used I-TASSER (that in CASP7 to CASP11 was in the top ranking [109, 110]. These two methods of structure prediction, operate in two different ways. I-TASSER [38], utilizes threading in order to predict the 3D structure of the proteins. It matches the protein sequences with similar folds to the target protein, even when no evolutionary relationship exists between the solved template and the target protein[51]. In AlphaFold, deep learning [44] is used, allowing the algorithm to create 3D structures even when model proteins are unavailable. Deep learning is a subset of machine learning techniques that uses a vast amounts of data, to learn and recognize patterns between different protein conformations without needing the researcher's intervention [44, 46]. Nevertheless, when comparing the two inferred 3D structures of the 132 interactors according to the TM-score [111], 65% of the proteins show the similar protein fold [106]. Thus, for the large majority of the ataxin-3 interactors the results obtained with the two 3D structures are expected to be similar.

In this chapter, it is addressed the effect of using the two 3D structure prediction methods for the 132 interactors reported in Sousa e Silva, et al. [106] by comparing the ataxin-3 interaction regions. Using this data main interaction regions were traced (molecular signatures) in ataxin-3. Furthermore, by using different 3D structure predictions and from the ataxin-3 interactors that exhibited similar binding pattern, it was possible to identify those proteins that present a different binding strength with the two ataxin-3 forms.

1. Materials and Methods

1.1 Docking and interaction predictions.

Using an in-silico methodology developed by Rocha, et al. [3], I predicted the interactions between ataxin-3 and its respective interactors studied by Sousa e Silva, et al. [106] using their 3D AlphaFold structure (available at <https://alphafold.ebi.ac.uk/>). Sousa e Silva et al. [106] excluded from there analyses ataxin-3 interactors that present six or more interactions at the polyQ tract(see **Supplementary Table 8** from Sousa e Silva et al. [106]) , since according to Rocha, et al. [6] these interactors lead to anomalous predictions, when studying ataxin-1. Here I address, if the proteins with six or more interactions in the ataxin-3 polyQ region behave differently from the remaining interactors.

Prior to submitting the protein structure to the docking server, the interacting residues (active and passive) were predicted using CPORT (<https://alcazar.science.uu.nl/services/CPORT/>) [82]. Once the residues were determined, docking experiments were performed for both ataxin-3 forms with their respective interactors, using HADDOCK webserver available at <https://bianca.science.uu.nl/haddock2.4/> [78, 79]. After the docking process was concluded, among the variety of complexes presented by HADDOCK, only the clusters with a z-score ≤ 0 were selected [78]. Out of the best four solutions for each cluster, only the structure with highest interactions, according to PDBePISA (<https://www.ebi.ac.uk/pdbe/pisa/>) [112], was selected for further data processing. These inferences were made using the PDBePISA docker image available at pegi3S (<https://pegi3s.github.io/dockerfiles/>), as well as the pisa_xml_extract script included in the utilities docker image, also available at pegi3S. This creates a file describing every individual interaction for each amino acid residue in the ataxin-3 protein, that can be opened using Excel.

1.2 Differences in 3D protein structures

TM-align [111], a web tool available at <https://zhanggroup.org/TM-align/>, was used to determine differences in protein structure between I-TASSER and AlphaFold. I consider that two proteins have the same protein fold if TM score is higher than 0.5.

1.3 Defining Binding regions

Since the JD (180 amino acids) is half of the wt-ataxin3 protein (361), we compared the number of interaction residues in these two regions, to classify the interactors as those that interact more at the JD and those that interact more at the C-terminal. Since the polyQ region is located in C-terminal region I performed these analyses with and without this region, and consider that an interactor has more interactions in its C-terminal region if the number of interactions reported in the C-terminal region is higher in both analyses. In further analyses the two groups are analysed separately.

To identify ataxin-3 interacting regions (IR) I use the same approach of Sousa e Silva, et al. [106] that take into account the amino acid positions that interact with more than 50% of the interactors and that are less than three consecutive amino acids apart from each other.

1.4 Statistical analysis

Statistical analyses were performed using SPSS, version 27.0 [105] using non-parametric tests. If the data is paired, we used a Sign test, while if the two samples are independent, a Mann-Whitney U test was used. The significance level is 5%.

2. Results and discussion

2.1 Ataxin-3 interactions using I-TASSER and AlphaFold 3D structures predictions are highly correlated

When considering the 3D AlphaFold and I-TASSER structure predictions of the 132 ataxin-3 interactors studied in Sousa e Silva et al. [106], the number of ataxin-3 interactions per site is larger when using Alphafold structures than when using I-TASSER structures (Sign test; $P < 0.05$; positive counts=242 versus negative counts =93). Nevertheless, the number of interactions along the ataxin-3 using the two methodologies is highly correlated (Pearson's correlation coefficient $R = 0.947$; $P < 0.00001$; $y = 1.1557x - 7.5366$, being x and y the occupancy frequency of each residue when using AlphaFold and I-TASSER, respectively, **Figure 13**). The number of interactions along

ataxin-3 using the two methodologies is still significantly correlated when considering the ataxin-3 interactors that show the same fold according to TM score (higher than 0.5; N=86, Sousa e Silva et al. [106]; Pearson's correlation coefficient $R = 0.95$) and when considering those interactors that show a different fold, according to TM score (lower than 0.5; N=45; Pearson's correlation coefficient $R = 0.89$). Therefore, the inferences made using I-TASSER could be the same as that made using AlphaFold 3D structures.

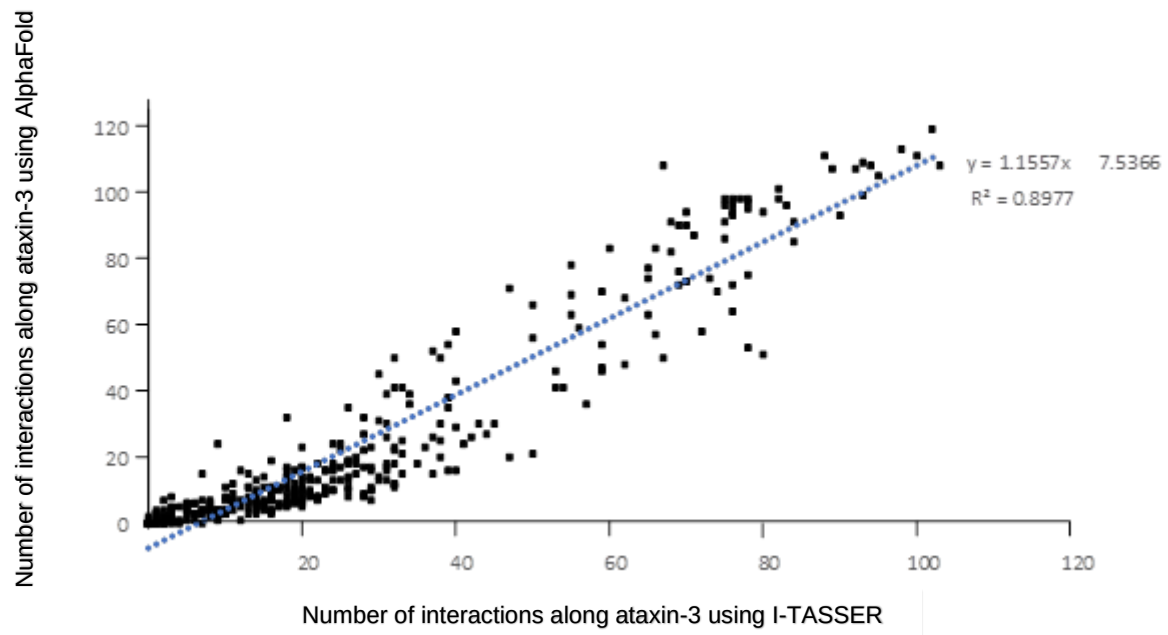


Figure 12. Pearson (R) linear correlation of the number of interactions along ataxin-3 when using 132 interactors, and the 3D protein structures prediction from AlphaFold and I-TASSER.

When looking at the interacting regions (defined by Sousa e Silva, et al. [106] as amino acid positions that interact with more than 50% of the interactors and that are less than three consecutive amino acids apart from each other), using the AlphaFold structures I identify six (located at 47-5; 85-90; 99-110; 125-129; 166-171; 175-191; **Figure 14**).

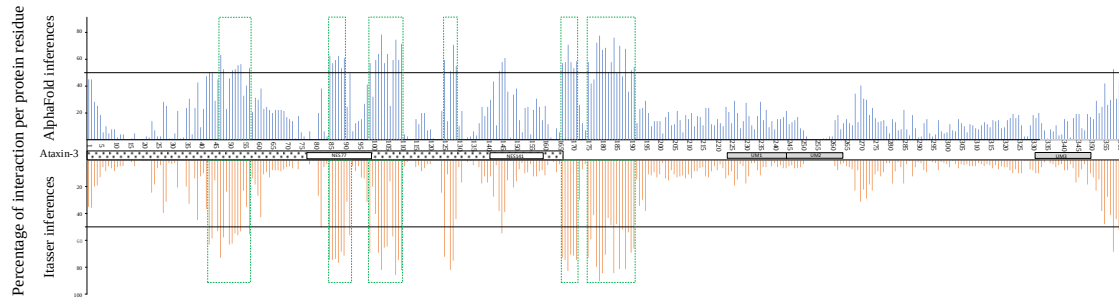


Figure 13. Percentage of interaction per protein residue at the ataxin-3 for 132 interactors using AlphaFold predicted 3D structures (upper panel, in blue) and I-TASSER 3D structures (lower panel, in orange). Green dotted boxes represent regions of interaction where more than 50 percent of the proteins show interaction with ataxin-3. The Josephin domain (JD) is assigned as a dotted box, while the NES77 and NES141 regions are marked with white boxes. UIM regions are also assigned within grey boxes

With the exception of region 125-129, all the others have been identified using I-TASSER structures for the ataxin-3 interactors. Therefore, the five interacting regions described in Sousa e Silva et al. [106] do not depend on the methodology used for 3D protein structure inference. For the five interaction regions in common when using AlphaFold and I-TASSER 3D predicted structures, the number of ataxin-3 interactions at each amino acid site is highly correlated (Pearson's correlation coefficient $R = 0.87$; $y = 1.2552x - 9.5268$, being x and y the occupancy of each residue when using AlphaFold and I-TASSER, respectively; **Figure 15**). These amino acids are thus important for ataxin-3 binding.

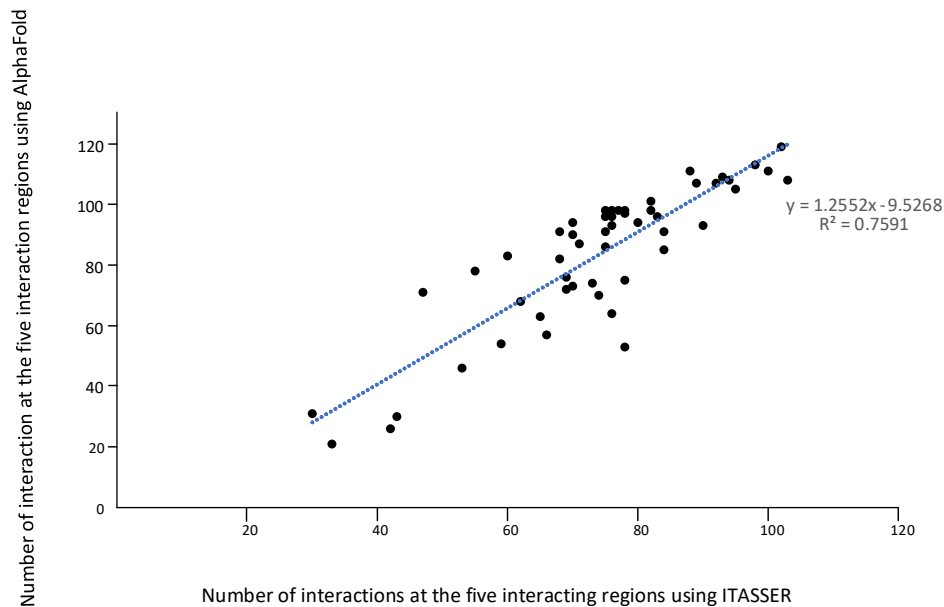


Figure 14. Pearson (R) linear correlation of the number of interactions considering the interacting regions (those marked with dotted green boxes in Fig. 14) in common between AlphaFold and I-TASSER 3D protein structures predictions.

2.2 Number of interactions in the JD and C-terminal regions are affected by the 3D protein structure used only in a small number of interactors

The increase in the number of interactions along ataxin-3 protein when considering AlphaFold structure predictions occurs when considering JD only (the first 180 amino acids; Sign test, $P < 0.05$, positive counts 96; negative counts 67) and the C-terminal region (from amino acid position 182 to 361; $P < 0.0001$; positive counts 146; negative counts 26), although is higher in the latter (**Figure 15**). This result suggests that proteins described as showing more interactions in JD than in the C-terminal region may be different in the two datasets. Using AlphaFold structure predictions, 104 interactors show a larger number of interactions in the JD regions, and 25 show a larger number of interactions in the C-terminal region, independently of considering the number of interactions in the polyQ region. As reported in Sousa e Silva, et al. [106], using I-TASSER structure predictions, 115 interactors show a larger number of interactions in the JD region, and only 15 show a larger number of interactions in the C-terminal region. For 100 interactors (94 that interact more at the JD and six that interact more at the C-terminal region; **Figure 16**) the results are the same independently of the protein structure prediction method used. As reported in the literature, a large set of ataxin-3 interactors bind mostly at the JD region (for references see **Table 1** in Sousa e Silva et al. [106]).

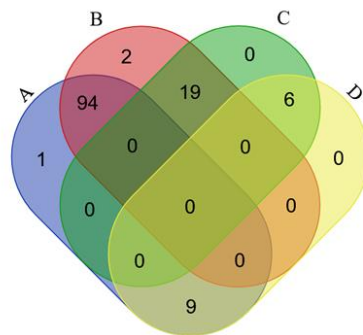


Figure 15. Overlap of the ataxin-3 interactors that show a larger number of interactions in the JD when considering AlphaFold (A) and I-TASSER (B) 3D structure predictions and those interactors that show a large number of interactions in the C-terminal region when considering AlphaFold (C) and I-TASSER (D) 3D structure predictions, independently of considering the number of interactions in the polyQ region (see **Supplementary Table 6** for Genelds).

Considering the proteins that mostly bind outside the ataxin-3 JD, the two proteins reported in the literature here studied (VCP (Gene ID 7415), and PCAF (Gene ID 8850)[113], when using I-TASSER protein structures the reported pattern was exhibited. However, when using AlphaFold protein structure predictions, the interactions for these proteins shown mainly interacting with the JD. Therefore, it seems that inferences made using I-TASSER could be more accurate than those made using AlphaFold. Since here, we only use for further analyses the proteins that show a similar pattern when using both

methodologies, and to obtain I-TASSER protein predictions implies a longer time (about 8 hours per protein), in the following work we use AlphaFold protein predictions, and for the proteins of interest, we then confirm the results using I-TASSER predictions.

2.3 Identification of the ataxin-3 interactors affected by the presence of an expanded polyQ

It has been hypothesized that when expanded, the ataxin-3 polyQ region would adopt a helical conformation, extending the preceding helix, and thus making the PPI interactions stronger [114-116]. Here I use the 100 proteins that present a similar binding behaviour using the two 3D protein prediction methodologies when considering the wt ataxin-3 form to address differences in binding strength between wt and exp ataxin-3. From these proteins, 12 (GeneID 3303, GeneID 4303, GeneID 5071, GeneID 6128, GeneID 6383, GeneID 7157, GeneID 9001, GeneID 9894, GeneID 10059, GeneID 57159, GeneID 83892, and GeneID 84676) were not analysed, since the in-silico methodology could not be performed using ataxin-3 exp form, and I-TASSER protein prediction structures [106]. From the 88 ataxin-3 interactors analysed using the two 3D structure predictions, 22 show more interactions with the exp ataxin-3 form than with the wt ataxin-3 form (**Table 2**).

Table 2. Interactors that show a higher number of interactions with the exp than the wt ataxin-3 forms using I-TASSER and AlphaFold 3D protein structures

GeneID	Number of interactions along the ataxin-3			
	<u>exp form</u> using AlphaFold	<u>wt form</u> using AlphaFold	<u>exp form</u> using I-TASSER	<u>wt form</u> using I-TASSER
310	86	78	56	52
396*	62	61	60	50
871	66	61	66	61
3105*	65	62	69	57
4702	77	73	75	72
4738	56	54	54	52
6175*	76	72	86	59
6390	95	78	55	52
7335*	63	61	73	62
10273	68	62	70	69
11284	80	80	66	66
23597*	89	78	89	73
29979	105	91	93	87
51170*	86	75	76	69
54708*	97	86	76	67
78990	53	51	56	55
119559*	76	74	93	77
121536	105	86	81	80
373863	86	70	76	74
387990	74	66	70	66
414919	81	71	61	60
493856	80	77	50	47

* Interactors that show more than 10% increase in the number of interaction residues in exp ataxin-3 relative to the wt form, reported in Sousa e Silva et al. [106]

This set of proteins is enriched in mitochondria membrane proteins according to PHANTER Go Slim Cellular component, a pattern that is not observed for the set (N=18) of proteins reported in Sousa e Silva et al. [106] as showing strong interaction with the expand ataxin-3 form not validated using AlphaFold protein structure inferences. PolyQ exp ataxin-3 is known to activate mitochondrial apoptotic pathway to induce neuronal death [117] thus supporting the predictive power of this approach.

Considering only those interactors with similar results using the two 3D structure predictions we validated 54% of the interactors that show different binding strength with the exp atxin-3 form when compared with the wt ataxin-3 form, previously identified using the I-TASSER protein structures, thus highlighting the need to perform the two methodologies to be confident that the results are not biased due to the methodology used.

2.4 Ataxin-3 interactors presenting six or more interactions in the polyQ region affect the results of the comparison of binding strength of ataxin-3 with exp and wt forms, when using I-TASSER and AlphaFold 3D structure predictions

When applying the in-silico methodology to ATAXIN-1 (ATXN-1=, Rocha et al. [3] observed that the number of interacting residues at the polyQ region was significantly associated with the wrong binding predictions, but such association was no longer significant when interactors showing six or more predicted interactions with the ATXN-1 polyQ region were removed. Therefore, in the work of Sousa e Silva et al. [106], the ataxin-3 interactors presenting six or more interactions at the polyQ region were not analysed, for both ataxin-3 forms. In this Chapter, I have analysed the 132 ataxin-3 interactors reported in Sousa e Silva et al.[106], independently of presenting six or more interactions when considering AlphaFold 3D structure predictions and the wt ataxin-3 form. When considering the 32 proteins that show a different binding pattern (towards JD or C-terminal region) using the different 3D protein structure predictions, only three interactors (GeneID 47, GeneID 836, and GeneID 9274) present six or more interaction at the wt polyQ region. Moreover, of the 100 ataxin-3 interactors that present similar binding pattern when using the two 3D protein structure predictions, only four (GeneID 9531, GeneID 54708, GeneID 90550, and GeneID 57159) present six or more interactions in the exp polyQ region. The distribution of these proteins in the two datasets is not different (Two-tailed Fisher exact test, $P > 0.05$). Therefore, the presence of six or more interactions in the polyQ region is not associated with behaving differently according to the 3D protein structure prediction.

When considering the exp polyQ region, the number of proteins that present six or more interactors is larger when considering I-TASSER (N=18) than when considering AlphaFold (N=3) 3D structure predictions. Only one (GeneID: 90550) is common in the two datasets. With the exceptions of GeneID: 54708 and GeneID: 5631 (that shows an increase and a decrease, respectively, in the number of interaction with the exp ataxin-3 form when using either structure prediction method) all the others (N=18) show opposite binding strength when comparing the binding with wt and exp forms and the two methodologies (17 show a large number of interactions when using I-TASSER than when using AlphaFold 3D structure predictions, and one show a larger number of interactions when using AlphaFold than when using I-TASSER 3D structure predictions). The number of interactors presenting less than six interactions at the exp polyQ tract, and that lead to different conclusions when comparing the wt and exp structures when using I-TASSER and AlphaFold is 35, while those that lead to similar conclusions is 33. Therefore, there is a significant effect (Two-tailed Fisher exact test; $P < 0.05$) of the number of interactions in the polyQ region in whether the predictions obtained with the two methodologies agree or not (2 and 18 vs 33 and 35).

Although the presence of six or more interactions in the polyQ region can affect the inferences on the relative binding strength with the ataxin-3 exp and wt forms, when using the 3D structure predictions obtained with the two methodologies, in the set of 22 ataxin-3 interactors that show agreement, there are nine proteins that show six or more interactions at the polyQ regions (**Figure 17**) when using I-TASSER structure predictions. Therefore, the presence of six or more interactions at the polyQ region does not necessarily imply a disagreement between the two approaches.

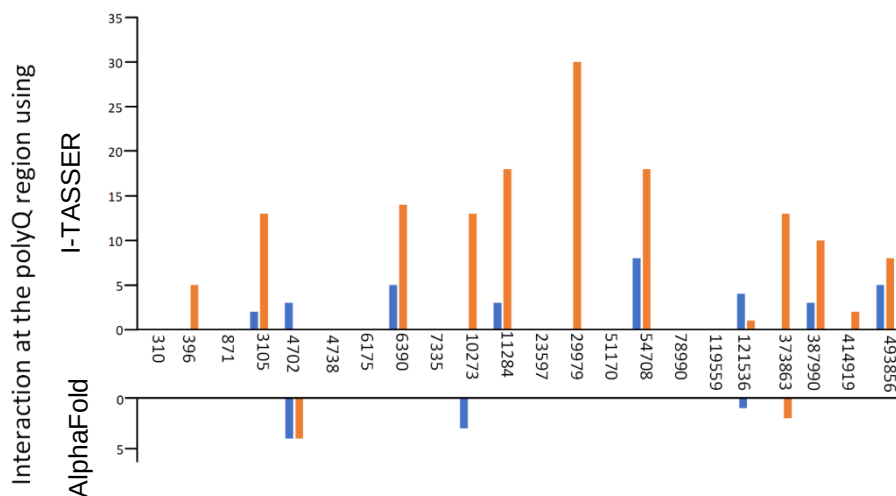


Figure 16. Number of interactions at the polyQ region using I-TASSER (upper panel) and AlphaFold (lower panel) for the ataxin-3 wt (in blue) and exp (in orange) form

Chapter 3: Using new interactors to evaluate their interacting pattern and predict ataxin-3 interaction regions

Throughout this work, the importance of the validation of the ataxin-3 protein network has been addressed, by using comparative approaches (as performed in Chapter 1) as well as the presence of interaction regions in both the wt and exp ataxin-3 forms (as performed in Chapter 2). In Chapter 2, I also addressed the impact of using two different methodologies, I-TASSER and AlphaFold, to obtain the 3D docking structure prediction for ataxin-3 with 132 interactors and infer the number and frequency of interactions at the wt ataxin-3 form. The number of ataxin-3 interacting sites that are inferred using AlphaFold and I-TASSER are correlated, and similar patterns of interaction with the ataxin-3 wt form were observed (**Figure 13**). Nevertheless, using AlphaFold 3D protein structure predictions there is an increase in the number of interactions with the wt ataxin-3 form, that causes differences in the classification of interactors as those that bind more at the Josephin domain (JD) and the C-terminal ataxin-3 regions. The two regions have similar sizes. For the 100 interactors that show more interactions at the JD using both methodologies, I identify those proteins that present different binding strength with ataxin-3 exp and wt forms, independently of the methodology used. As reported in Sousa e Silva, et al [106], expansion of the polyQ tract does not equally affect all interactors, and those with increased binding strength to the expanded form may be linked to the onset of the SCA3 disease [2, 3, 106].

In this chapter, the newly disclosed interactors selected in Chapter 1 (N=99) were analyzed using 3D structure predictions obtained with AlphaFold, to evaluate if their interaction behavior is similar to that observed for true ataxin-3 interactors (see previous Chapter), thus, validating these proteins as ataxin-3 interactors. By comparing the binding strength with the wt and exp ataxin-3 forms, I identify further putative SCA3 related proteins. For the ataxin-3 interactors that present such pattern I validate the inferences using I-TASSER 3D predicted structures. It should be noted that while AlphaFold predicted structures are available for download at AlphaFoldDB for all proteins in UniProt, I-TASSER predictions must be obtained using the I-TASSER server, that, on average, takes eight hours per interactor.

1. Material and Methods

1.1 Comparison between the interactors of the previous dataset and the new dataset

In Chapter 1 (**Supplementary Table 3**), 99 new interactors were selected among the reported set of ataxin-3 interactors present in both the main EvoPPI3 and the polyQ_22 databases. After obtaining their AlphaFold predicted structures, as described in Chapter 2, their active and passive residues were obtained using CPORT (<https://alcazar.science.uu.nl/services/CPORT/>) [82]. Since PDB files obtained from AlphaFold contained headings that interfere in CPORT analysis, for each file they have been removed manually. These PDB files were then submitted to a similar docking process as the one described in Materials and Methods in Chapter 2, with both ataxin-3 wt and exp forms. Due to limitations in the HADDOCK methodology, 5 proteins were not analyzed (GeneID 1778, GeneID 3190, GeneID 3309, GeneID 7184, GeneID 10594).

The selection of the best structures cluster after the docking, followed the same methodology described in Chapter 2 and the data was processed by PDBePISA in order to describe every single interaction per ataxin-3 protein residues within a excel spreadsheet.

For comparative analyses I used the 100 proteins that strongly bound to the JD described in Chapter 2.

1.2 Detecting different behavior between proteins with different evidence for being ataxin-3 interactors

The proteins reported in the **Supplementary Table 3**, show different support for being ataxin-3 interactors, since not all are reported in all EvoPPI3 databases. With this in consideration, here I divide the ataxin-3 interactors that strongly bind to the JD in two groups, those that have been reported in *D. melanogaster* modifiers and *D. rerio* polyQ databases, and therefore have multiple evidence to support the claim that they are ataxin-3 interactors, and those that are present in *D. melanogaster* modifiers or *D. rerio* polyQ databases for which support as ataxin-3 interactors is not so strong.

1.3 Tracing the proteins that bind strongly to the exp ataxin-3 than to the wt ataxin-3 form

To identify the proteins that bind differently with the exp ataxin-3 form, the methodology described above for the wt ataxin-3, has been applied to the exp ataxin-3 form.

For the proteins that present an increase of 10% in binding strength with exp ataxin-3 compared with the wt ataxin-3 form, additional protein docking has been performed using

the I-TASSER 3D structure predictions. These results were then compared with those obtained using AlphaFold-based approach and the interactors that displayed consistent results in both methodologies, were considered. These proteins were then merged with the 22 proteins identified in Chapter 2, to address their enrichment in Cellular components, using PANTHER Gene Ontology analysis [60].

2. Results and Discussion

2.1 Analyzing possible differences in the interaction patterns of ataxin-3 wt when using the new selected interactors

Of the 94 ataxin-3 interactors analyzed 88% (83 interactors) show a large number of interactions in the JD than in the C-terminal region (**Table 3; see Supplementary Table 7** for the ataxin-3 interactors that strongly bind to the ataxin-3 C-terminal region). The observation that there is a significant amount of proteins that bind primarily to the JD, is supported by the literature (see **Table 1** from Sousa e Silva, et al [106]).

Table 3. Proteins that mainly bind to the JD

<i>Difference of interactions between JD and the C-terminal region</i>			
<i>Uniprot ID</i>	<i>GeneID</i>	<i>JD - C-terminal</i>	<i>JD - C-terminal(without polyQ)</i>
<i>J3KSQ2</i>	<i>1213</i>	21	21
<i>O00161</i>	<i>8773</i>	28	28
<i>O00232</i>	<i>5718</i>	27	27
<i>O00487</i>	<i>10213</i>	14	14
<i>O43707</i>	<i>81</i>	16	16
<i>O60884</i>	<i>10294</i>	12	12
<i>O75400</i>	<i>55660</i>	17	15
<i>O95373</i>	<i>10527</i>	32	32
<i>P04075</i>	<i>226</i>	38	38
<i>P04406</i>	<i>2597</i>	34	34
<i>P04792</i>	<i>3315</i>	42	42
<i>P04843</i>	<i>6184</i>	10	9
<i>P07814</i>	<i>2058</i>	21	21
<i>P11498</i>	<i>5091</i>	2	2
<i>P12268</i>	<i>3615</i>	15	15
<i>P12814</i>	<i>87</i>	39	39
<i>P13473</i>	<i>3920</i>	6	3
<i>P13639</i>	<i>1938</i>	30	30
<i>P14868</i>	<i>1615</i>	7	5
<i>P17980</i>	<i>5702</i>	13	13
<i>P17987</i>	<i>6950</i>	24	24
<i>P22234</i>	<i>10606</i>	24	24
<i>P25789</i>	<i>5685</i>	15	15
<i>P26196</i>	<i>1656</i>	16	16
<i>P27694</i>	<i>6117</i>	40	40
<i>P28066</i>	<i>5686</i>	12	12

Table 3 (continuation). Proteins that mainly bind to the JD

<i>Diference of interactions between JD and the C-terminal region</i>			
<i>Uniprot ID</i>	<i>GeneID</i>	<i>JD - C-terminal</i>	<i>JD - C-terminal(without polyQ)</i>
<i>P28074</i>	<i>5693</i>	28	28
<i>P31948</i>	<i>10963</i>	17	17
<i>P35998</i>	<i>5701</i>	11	11
<i>P43686</i>	<i>5704</i>	28	28
<i>P48444</i>	<i>372</i>	39	39
<i>P49591</i>	<i>6301</i>	31	31
<i>P49721</i>	<i>5690</i>	32	32
<i>P50990</i>	<i>10694</i>	33	33
<i>P53621</i>	<i>1314</i>	38	38
<i>P61313</i>	<i>6138</i>	14	14
<i>P61619</i>	<i>29927</i>	42	42
<i>P62249</i>	<i>6217</i>	16	16
<i>P62701</i>	<i>6191</i>	13	13
<i>P62753</i>	<i>6194</i>	27	27
<i>P62826</i>	<i>5901</i>	31	31
<i>P78371</i>	<i>10576</i>	22	22
<i>Q00839</i>	<i>3192</i>	35	35
<i>Q04637</i>	<i>1981</i>	18	18
<i>Q15029</i>	<i>9343</i>	31	31
<i>Q15046</i>	<i>3735</i>	41	41
<i>Q15393</i>	<i>23450</i>	26	26
<i>Q16363</i>	<i>3910</i>	8	8
<i>Q16531</i>	<i>1642</i>	27	27
<i>Q59ES8</i>	<i>4670</i>	3	3
<i>Q7L576</i>	<i>23191</i>	28	28
<i>Q7Z6Z7</i>	<i>10075</i>	27	27
<i>Q8WWM7</i>	<i>11273</i>	13	13
<i>Q93009</i>	<i>7874</i>	1	1
<i>Q96CS3</i>	<i>23197</i>	37	37
<i>Q99460</i>	<i>5707</i>	31	31
<i>Q99615</i>	<i>7266</i>	14	14
<i>Q99873</i>	<i>3276</i>	40	40
<i>Q9P2J5</i>	<i>51520</i>	36	36
<i>Q9UNM6</i>	<i>5719</i>	26	26
<i>Q9Y3B4</i>	<i>51639</i>	13	13
<i>O14980</i>	<i>7514</i>	36	36
<i>O43242</i>	<i>5709</i>	25	25
<i>P04350</i>	<i>10382</i>	32	32
<i>P08238</i>	<i>3326</i>	2	2
<i>P0DMV9</i>	<i>3304</i>	16	15
<i>P14618</i>	<i>5315</i>	21	21
<i>P20618</i>	<i>5689</i>	16	11
<i>P23528</i>	<i>1072</i>	29	29

Table 3 (continuation). Proteins that mainly bind to the JD

<i>Diference of interactions between JD and the C-terminal region</i>			
<i>Uniprot ID</i>	<i>GeneID</i>	<i>JD - C-terminal</i>	<i>JD - C-terminal(without polyQ)</i>
P24539	515	2	2
P26640	7407	42	42
P34931	3305	38	38
P38646	3313	15	15
P60900	5687	26	26
P62258	7531	3	3
P62333	5706	13	13
P68371	10383	19	19
Q3ZCM7	347688	40	40
Q99832	10574	16	11
Q9Y617	29968	33	33
P17066	3310	29	29
P00558	5230	5	3
Q13200	5708	7	4

For these ataxin-3 interactors, the interacting regions, as described in Chapter 2, were obtained (**Figure 18**; upper panel). Although the frequency of interactions at the ataxin-3 wt form vary when comparing the 83 ataxin-3 interactors here studied (**Figure 18**; upper panel) and the 100 reported in Chapter 2 (**Figure 18**; lower panel; Sign test; $P < 0.00001$; $y = 0.9074x + 8.1189$, being x and y the occupancy frequency of each residue when using AlphaFold 3D structure predictions of 83 ataxin-3 interactors analysed in this chapter and those 100 of chapter 2, respectively, **Figure 19**).

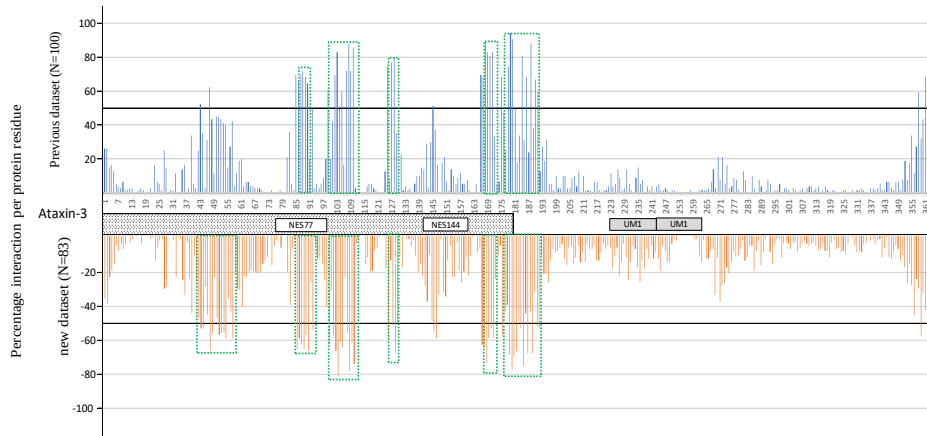


Figure 17. Comparison between the interaction percentages of each amino acid for ataxin-3 wt form, for both the new dataset (upper panel, blue color) and the previously studied dataset (lower pannel, orange color). In green dotted boxes are defined the interaction regions

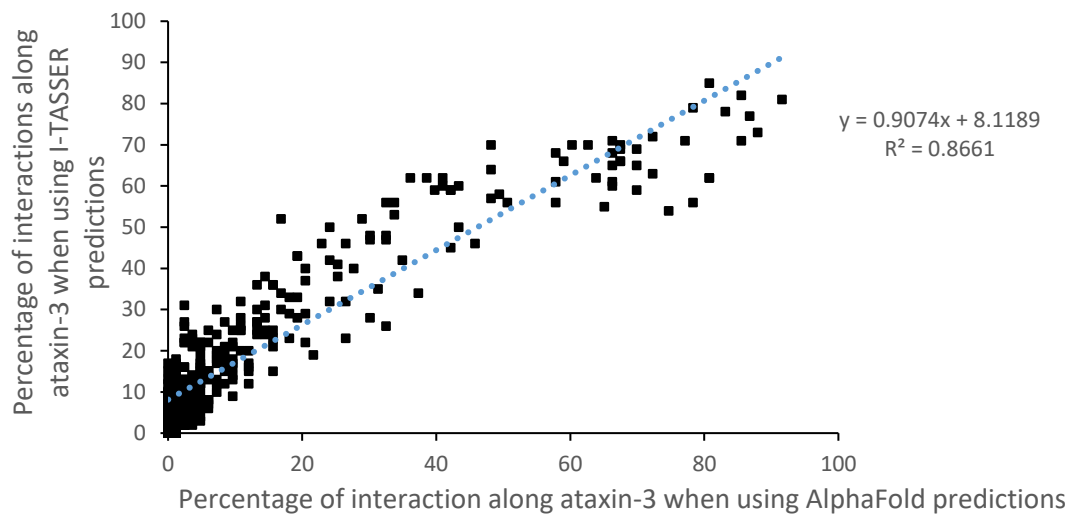


Figure 18. Pearson (R) linear correlation of the percentage of interactions along ataxin-3 when using 3D protein structures prediction from AlphaFold for 83 ataxin-3 interactors here analysed and 100 ataxin-3 interactors analysed in Chapter 2, that have in common higher frequency binding with the JD region

When looking at the interacting regions (defined by Sousa e Silva, et al. [106] as amino acid positions that interact with more than 50% of the interactors and that are less than three consecutive amino acids apart from each other), using the 83 AlphaFold structures here studied, I identified five interacting regions, located at ataxin-3 amino acid regions 85-90, 99-110, 125-129, 166-171, and 175-191 (**Figure 18**). These interacting regions, with the exception of the region at amino acid positions 42-57, are the same obtained using the 100 ataxin-3 interactors analysed in Chapter 2. Comparison of the interaction percentages between these regions in the two datasets show, however not significant differences (Mann-Whitney test; $P > 0.05$; **Supplementary Table 8**).

The absence of a binding region at amino acid positions 42-57, present in the proteins analysed in Chapter 2, could be related with differences in protein function of the two datasets. Indeed, the two groups of ataxin-3 interactors are different according to PantherDB class of proteins (**Table 4**). The dataset that contains the new interactors, is enriched in translational proteins, the dataset used in Chapter 2 is enriched in ubiquitin-protein ligase. Both protein classes have been reported as ataxin-3 interactors since ataxin-3 is involved in mRNA translation [118] and ubiquitin related systems [4, 119]. Further analyses are, however, needed to associate the differences in the interaction regions to the function of the proteins.

Table 4. Comparison between the two datasets protein classes

Protein class	N=100	N=94	Fold Enrichment	P value
Translational protein	2	14	8,52	3,85E-04
Ubiquitin-protein ligase	17	1	0,07	2,54E-04

2.2 Detecting different behavior between proteins with different evidence for being ataxin-3 interactors

Among the previously selected 83 interactors, it is important to note that these interactors present varying degrees of supporting evidence for being ataxin-3 interactors. This distinction allows for the formation of two groups. Group 1 comprising 39 proteins with multiple evidence for being ataxin-3 interactors since they are present in both *D. melanogaster* modifiers and *D. rerio* polyQ_22 databases (N=39; **Supplementary Table 9; Supplementary Table 10**) and the group 2, contains 44 the proteins with less supporting evidence, as they are present in only one of the two databases (N=44) .

The percentage of interactions along the wt ataxin-3 in the both groups is different (Mann-Whitney test; $P > 0.05$; group 1 does not contain higher interaction percentages than the group 2). Additionally, the percentage of interactions between both groups is highly correlated (Pearson's correlation coefficient $R = 0.964$; $P < 0.00001$; $y = 0,9074x - 0.81189$, being x and y the percentage of interaction of each residue in group 1 and group 2, respectively; **Figure 19**). Moreover, when analyzing the interaction regions between group 1 and group 2, they are the same except for the region at amino acid positions 42-57, present in group 1 (**Figure 20**). This interaction region is also observed when considering the 100 interactors analyzed in Chapter 2. The two groups are significant different, according to PANTHER GO-Slim Molecular Function, being the revealed enrichment in proteins responsible for RNA binding in group 2 of proteins 8 (**Table 5**). 50% of the proteins associated with the enrichment in RNA binding of group 2 are those associated with translational proteins when analysing the two groups together, thus suggesting that function of these proteins could be associated with the first interacting region at the ataxin-3 amino acid positions 42-57.

In conclusion, the 83 proteins here analyzed seem to be ataxin-3 interacting proteins, and thus, all can be used in the following analysis.

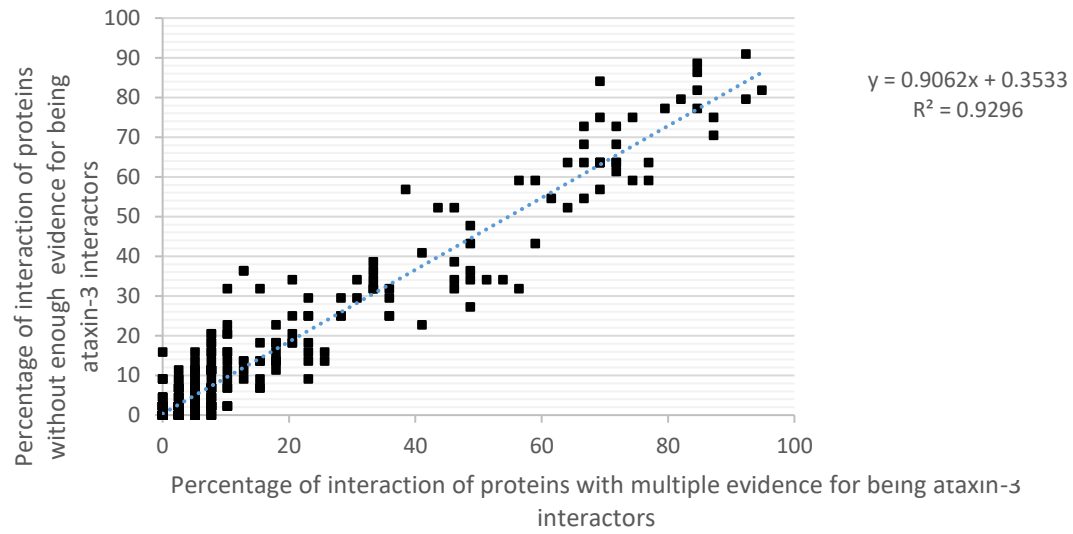


Figure 19. Pearson (R) linear correlation of the percentage of interactions along ataxin-3 of the interactors with more evidence for being ataxin-3 interactors and less evidence for being ataxin-3 interactors

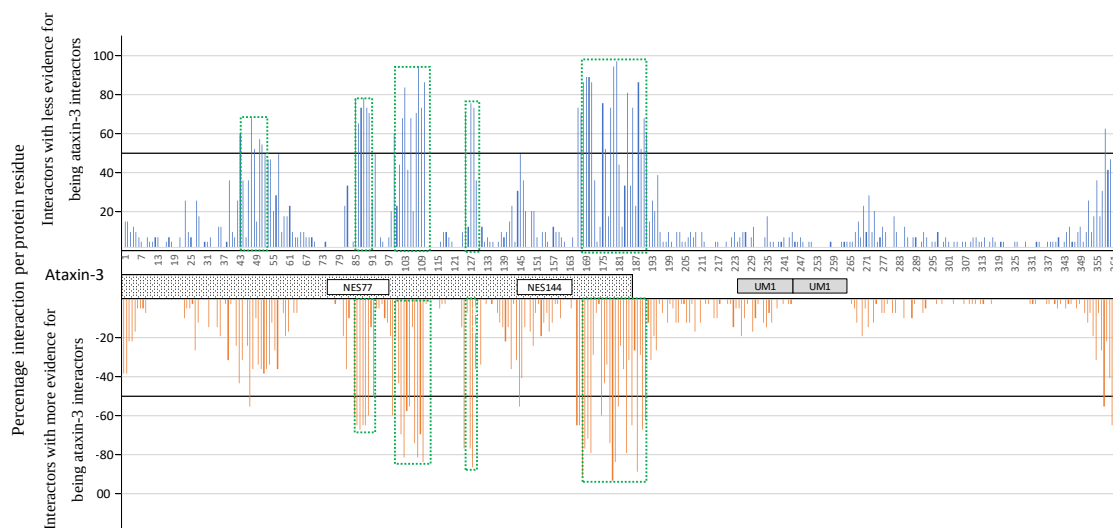


Figure 20. Comparison between the interaction percentages of each amino acid for ataxin-3 wt form, for both the dataset of interactors with less evidence for being ataxin- interactors (upper panel, blue color) and the dataset of interactors with more evidence for being ataxin-3 interactors (lower pannel, orange color). In green dotted boxes are defined the interaction regions

Table 5.Comparison between the two datasets Molecular functions

PANTHER GO-Slim Molecular Function	N=44	N=39	Fold Enrichment	P value
RNA binding	14	1	0,08	3,55E-04

2.3 Detecting the proteins that are more prone to bind to the EXP ataxin-3 protein

As observed in Sousa e Silva, et al [106], the presence of an expanded polyQ region may affect the binding strength of the interactions differently among the ataxin-3 interactors. The proteins that suffer from these variations, may suffer an increase in the number of interactions with the exp ataxin-3 relative to the wt and this may be an indicator that those proteins present an important role in the SCA3 disease. Therefore, an analysis of the 83 interactors was conducted with the objective to evaluate which of the proteins present these variations and that, are possible related to SCA3 disease.

To test the hypothesis, the total number of interactions presented by each protein was analyzed (**Supplementary Table 12**). 45 interactors shown an increase in the number of interactions with the exp ataxin-3 relative to the wt. Due to time constrains, for the following analysis only the ataxin-3 interactors that exhibited an increase larger than 10% with the exp ataxin-3 form when compared with the wt ataxin-3 form were selected (N=24; **Figure 17**). Moreover, employing a conservative approach, the three proteins (GeneID 3326, GeneID 5706, GeneID 6194) that change status according to the inclusion of the polyQ were also removed (**Table 6**).

Table 6. Proteins that show an increase in their interactions relative to the ataxin-3 wt when interacting with the exp form (in grey are those proteins that have been removed because they presented more than 6 interactions in the polyQ region; the values inside the () represent the number of ataxin-3 interactions excluding the polyQ region)

<i>UniprotID</i>	<i>GeneID</i>	Number of ataxin-3 interactions between expanded and wild type forms	Percentage of interactions considering the polyQ region (%)	Percentage of interactions excluding the polyQ region (%)
<i>O00161</i>	<i>8773</i>	7 (7)	19	19
<i>O00232</i>	<i>5718</i>	6 (6)	14	14
<i>O00487</i>	<i>10213</i>	14 (14)	30	30
<i>P12814</i>	<i>87</i>	9 (9)	21	21
<i>P13473</i>	<i>3920</i>	7 (4)	13	8
<i>P14868</i>	<i>1615</i>	11 (12)	23	27
<i>P17980</i>	<i>5702</i>	14 (6)	30	13
<i>P17987</i>	<i>6950</i>	7 (7)	13	13
<i>P31948</i>	<i>10963</i>	12 (7)	28	16
<i>P35998</i>	<i>5701</i>	6 (6)	12	12
<i>P43686</i>	<i>5704</i>	16 (16)	36	36
<i>P49721</i>	<i>5690</i>	8 (8)	18	18
<i>P53621</i>	<i>1314</i>	12 (12)	23	23
<i>P61313</i>	<i>6138</i>	20 (17)	43	37
<i>P62701</i>	<i>6191</i>	8 (8)	18	18
<i>P62753</i>	<i>6194</i>	10 (-1)	26	-3
<i>Q15393</i>	<i>23450</i>	10 (10)	19	19
<i>Q7Z6Z7</i>	<i>10075</i>	9 (6)	20	13
<i>Q99460</i>	<i>5707</i>	17 (12)	38	27
<i>O14980</i>	<i>7514</i>	16 (1)	21	1
<i>P08238</i>	<i>3326</i>	9 (-2)	20	-5
<i>P24539</i>	<i>515</i>	26 (26)	54	54
<i>P62333</i>	<i>5706</i>	7 (-1)	12	-2
<i>P68371</i>	<i>10383</i>	6 (6)	14	14

For 20 ataxin-3 interactors I validate this inference using I-TASSER 3D structure predictions (GeneID:10075 has not been analyzed do to size limitation of this methodology. Using HADDOCK as a docking method, I performed the docking of these proteins with both exp and wt ataxin-3 I-TASSER structures . The results of this analysis revealed that among these 20 interactors only seven shown an increase in the number of interactions when comparing exp and wt ataxin-3 forms independently of the methodology used (**Table 7** and **Table 8**).

Table 7. Proteins that show an increase in ataxin-3 exp interactions higher than 10% when using the I-TASSER ataxin-3 predicted structure with I-TASSER interactors predicted structures (In grey are located the proteins whose results are in agreement to the ones presented in the IT-AF docking; the values inside the **()** represent the number of ataxin-3 interactions excluding the polyQ region)

<i>UniprotID</i>	<i>GeneID</i>	Number of ataxin-3 interactions between expanded and wild type forms	<i>Percentage of interactions considering the polyQ region (%)</i>	<i>Percentage of interactions excluding the polyQ region (%)</i>
O00161	8773	10 (11)	14	16
O00232	5718	-7 (-20)	-8	-24
O00487	10213	-3 (-22)	-3	-26
O14980	7514	27 (12)	36	17
P12814	87	4 (-5)	6	-7
P13473	3920	5 (-10)	9	-19
P14868	1615	-22 (-21)	-26	-25
P17980	5702	-10 (-4)	-12	-5
P17987	6950	0 (-1)	0	-1
P24539	515	6 (0)	13	0
P31948	10963	-10 (-10)	-13	-13
P35998	5701	-2 (-7)	-2	-8
P49721	5690	0 (0)	0	0
P53621	1314	27 (17)	64	44
P61313	6138	6 (-2)	8	-3
P62701	6191	7 (7)	11	11
P68371	10383	6 (-13)	10	-22
Q99460	5707	10 (2)	12	2
P43686	5704	22 (18)	31	28
Q15393	23450	-4 (-3)	-5	-4

Table 8. Proteins that presented an increase in ataxin-3 exp independently of the methodology used (ataxin-3 I-TASSER structure vs interactors I-TASSER structure; ataxin-3 I-TASSER structure vs interactors AlphaFold structures)

Ataxin-3 interactors (Gene synonyms; GeneID; UniprotID)	Methodology used to support the interaction
SNAP23; 8773; O00161	<i>Mass spectrometry [19]</i>
XPO1; 7514; O14980	<i>Mass spectrometry [19]</i>
ATP5F1; 515; P24539	<i>Mass spectrometry [19]</i>
RPS4X; 6191; P62701	<i>Mass spectrometry [19]</i>
PSMD1; 5707; Q99460	<i>Mass spectrometry [19]</i>
PSMC4; 5704; P43686	<i>Mass spectrometry [19]</i>
COPA; 1314; P53621	<i>Mass spectrometry [19]</i>

The protein synaptosome associated protein 25 (SNAP25; GeneID 8773) is involved in the formation, recycling, and transport of synaptic vesicles [120]. Its distribution in brain is tightly regulated [120] and thus changes in the among of this protein in the different regions of the brain could contribute for ataxia. Indeed mutations in *Snap25* have been reported as causing ataxia in mice [121, 122]. Moreover, Snap 25 in conjugation with DnaJC5 protein (a cysteine string protein- α), is involved in the release of neurodegenerative proteins through a non-canonical exocytosis pathway, a mechanism for the exit of aggregate-prone proteins in the cell [123]. Therefore, the strong binding of ataxin-3 with this protein could change the levels of the protein in the brain, as well as the mechanism of exocytosis.

Exportin-1 (Xpo1; Gene ID 7514) is known to regulate the nucleocytoplasmic distribution of exp polyQ proteins such ataxin-3 [124]. In *Xpo1* knockdown mice, the exp polyQ form is retained in the nucleus but not the wt form, and thus suggesting that the exp polyQ form possesses nuclear export activity. Nuclear export of expanded polyQ protein helps in maintaining the cellular heat shock response during the course of polyQ pathogenesis. This protein shows a temporal decline in both wild-type and polyQ transgenic mice, that could account for the appearance of the disease late in life.

ATP synthase peripheral stalk-membrane subunit b (ATP5PB; GeneID 515) has been reported as showing altered interactions with mutant ataxin-3 vs normal ataxin-3 in the Ki150 mice model [125].

Our results also suggest that the exp ataxin-3 shows a tendency to interact with subunits of protein complexes such those of the proteasome such as the proteasome 26S subunit, non-ATPase 1 (PSMD1/RPN2; Gene ID 5707) and proteasome 26S subunit ATPase 4 (PSMC4/RPN4; Gene ID 5704). As suggested by Piasecki et al. [125], a disrupted

interaction of mutant ataxin-3 with subunits of complexes with circular structure such the ones here found, might lead to increase in formation of self ataxin-3 annular aggregates.

These 7 proteins were the combined to the group of 22 proteins from **Table 2** that also show a large number of interactions when comparing the exp and wt ataxin-3 forms. This dataset still showed an enrichment in mitochondria membrane proteins, as reported in Chapter 2, according to PANTHER Go Slim Cellular component, and also in proteins related to proteasome regulation. These results that are in agreement with the literature, since ataxin-3 interacts with proteins of the ubiquitin-protease system [120]. Furthermore, a large majority of this set (N=17) was enriched in proteins located in the cytoplasm showing that ataxin-3 interactions and poly-Q induced pathogenesis seem to occur mainly in the cytoplasm [11, 31].

Table 9. TM scores for the proteins when considering the different 3D structure prediction methods (in grey are the proteins that presented an increase in the number of interactions with the ataxin-3 exp form when compared with the wt form, independently of the methodology used)

<i>UniprotID</i>	<i>GeneID</i>	TM score
<i>O00161</i>	<i>8773</i>	0.40515
<i>O00232</i>	<i>5718</i>	0.88517
<i>O00487</i>	<i>10213</i>	0.84082
<i>P12814</i>	<i>87</i>	0.29548
<i>P13473</i>	<i>3920</i>	0.22558
<i>P14868</i>	<i>1615</i>	0.94116
<i>P17980</i>	<i>5702</i>	0.81317
<i>P17987</i>	<i>6950</i>	0.96488
<i>P31948</i>	<i>10963</i>	0.24051
<i>P35998</i>	<i>5701</i>	0.70288
<i>P43686</i>	<i>5704</i>	0.71788
<i>P49721</i>	<i>5690</i>	0.9864
<i>P53621</i>	<i>1314</i>	0.42545
<i>P61313</i>	<i>6138</i>	0.96923
<i>P62701</i>	<i>6191</i>	0.99016
<i>Q15393</i>	<i>23450</i>	0.97996
<i>Q99460</i>	<i>5707</i>	0.89738
<i>O14980</i>	<i>7514</i>	0.95527
<i>P24539</i>	<i>515</i>	0.65166
<i>P68371</i>	<i>10383</i>	0.97792

Additionally, it is possible to observe that among these 7 selected proteins, only two of them exhibited a TM score below the threshold of 0.5 (**Table 9**). This shows us that the obtained results are independent of the structural differences present when using the two different 3D structures prediction methods.

In conclusion, according to the hypothesis that the expanded polyQ region affects PPIs, I find evidence that a small set of ataxin-3 interactors (31 out of 228) are predicted to be affected by the expanded polyQ tract. These PPIs are expected to have an important role in the SCA3 pathogenesis. Experimental evidence is however needed to validate these predictions. Since most of these interactions have been reported as modifier genes in mutant ataxin-3 *D. melanogaster* ataxin-3, as reported in Sousa e Silva, et al. [106], this model system seems to be appropriate to address the role of these proteins in SCA3 pathology.

Chapter 4: Using an AlphaFold ataxin-3 predicted structure for a docking simulation

High-resolution experimental structures of a protein can be obtained using techniques such as X-ray crystallography, cryo-EM, and nuclear magnetic resonance (NMR). Despite substantial technological advancements, structures available at the Protein Data Bank (PDB) are still limited. This is because these methods are costly and time consuming, but also due to difficulties associated with protein expression and purification [33]. For larger proteins with extensive disordered regions such as ataxin-3 it is not possible to fully uncover the full structures [32, 126, 127]. Therefore, the experimental information available for the 3D structures of ataxin-3, obtained using NMR technique, are for the isolated Josephin domain [17, 126, 128] and those of the ubiquitin-interacting motif (UIM) domains located between the Josephin domain and the poly-Q segment [129]. The different structures of the ataxin-3 JD represent different conformational states, suggesting that the JD is in a dynamic equilibrium between open and closed conformations [128].

In chapter 2, and 3 I used the 3D ataxin-3 structure for the wt and exp forms predicted by I-TASSER, since these predictions show a good agreement with the secondary structure patterns of the PDB structures [130]. The I-TASSER ataxin-3 models are dominated by helices and coils and only a small proportion of the residues form β -structures [130]. I-TASSER prediction models for the exp form have more helical conformations and less β -structures and coils than the models for the wt form [130].

The presence of different JD structures suggest that ataxin-3 do not fold into a single conformation, but instead its able to 'flip' between different conformations to be functional. It is possible that alternative conformations interact differently with the network of protein interactors. In this chapter I use ataxin-3 AlphaFold 3D structure prediction to address how these models agree with the PDB structures and how the structure prediction used can affect inferences on PPIs, by comparing the interactions of both 3D structures with the same set of interactors.

1. Materials and methods

1.1. Addressing the similarities between both IT and AF predicted ataxin-3 structures with different the experimental ataxin-3 structures

To analyze structure difference of I-TASSER and AlphaFold 3D structures predictions, as well as structures determined experimentally for the ataxin-3 JD available at PDB database [131] (available at: <https://www.rcsb.org>) , I used TM align [111].

1.2. Docking of the interactors and evaluations of the interactions

To evaluate how the interactions may vary with different ataxin-3 conformations, I used AlphaFold 3D structures for both ataxin-3 and 30 interactors reported in Chapter 2, for which experimental evidence exist to support their interaction with the JD and the C-terminal region (see **Table 1** in Silva e Sousa et al. [106]. As in Chapter 3, active and passive sites were determined for the wt ataxin-3 AlphaFold 3D structure prediction, using CPORT. Docking was preformed using HADDOCK. The following steps of the analysis follow the same methodology detailed in Chapter 2 and 3.

2. Results and Discussion

2.1. I-TASSER and AlphaFold 3D structure predictions show similarity with different JD PDB structures

TM-score for AlphaFold and I-TASSER ataxin-3 structure predictions is 0.44232. This value is below the threshold of 0.5 defined by Zhang, et al [111], which shows significant differences in their structure. To find support for the two ataxin-3 structure predictions I compared these structures with experimental JD structures available at PDB (**Table 10**). I-TASSER predicted structure is supported by the JD 2AGA [17] structure, since the two show a TM score of 0.96860. AlphaFold structure prediction is also supported by the 2JRI [132] structure (TM score = 0.85611). Therefore, the two different ataxin-3 conformations may exist.

Table 10. TM score calculation between the ataxin-3 AlphaFold and I-TASSER predicted structures with ataxin-3 structures available at PDB database

Ataxin-3 from PDB database (Identity ID)	Ataxin-3 predicted structure	TM Score
2KLZ	I-TASSER	0.49424
	AlphaFold	0.52515
2AGA	I-TASSER	0.96860
	AlphaFold	0.69168
1YZB	I-TASSER	0.39708
	AlphaFold	0.4601
2JRI	I-TASSER	0.70097
	AlphaFold	0.85611

2.2. The use of the Ataxin-3 AlphaFold structure suggests that the majority of the interactors interact more at the Ataxin-3 C-terminal region than at the JD, in contrast with experimental evidence

Of the 30 ataxin-3 interactors that, when using either AlphaFold or I-TASSER structures, and the ataxin-3 I-TASSER structure, show more interaction sites at the JD than at the C-terminal region, only 10 proteins show this pattern when using the ataxin-3 AlphaFold protein structure prediction (GeneID 41, GeneID 293, GeneID 871, GeneID 7846, GeneID 8841, GeneID 119559, GeneID 203068, GeneID 396, GeneID 824, GeneID 1019). These findings are in contrast with experimental data [106], since the JD is reported as the main binding site in ataxin-3 interactions.

In summary, the use of the ataxin-3 AlphaFold structure prediction yielded interactions that significantly differ from those described in the literature, that are in agreement with those obtained when using I-TASSER structure prediction, suggesting that AlphaFold ataxin-3 conformation may be more transient than that predicted using I-TASSER, and thus, in future work the ataxin-3 I-TASSER structure prediction should be used.

Conclusion and Future work

In this work, 226 ataxin-3 interactors were used to perform *in-silico* analyses in order to find evidence that these proteins interact with ataxin-3, define ataxin-3 interacting regions, and identify those interactors that present different interaction strength with the exp ataxin-3 compared with the wt ataxin-3 forms. From this work, I traced hallmarks that allow to accurately predict and confirm ataxin-3 interactors. Here I also address the effect of using different 3D structure predictions of both the ataxin-3 interactors and the ataxin-3 wt form.

Ataxin-3 interactors have been identified using multiple experimental techniques such as yeast two-hybrid (Y2H) and mass spectrometry. However, such techniques, may produce a high rate of false positives, thus being difficult to identify the true ones [6]. For this reason, in Chapter 1, multiple cross analyses between the different databases available at EvoPPI3 [58], such as main databases, polyQ databases, predicted interactomes and modifier genes from model species) were performed. With these analyses, it was possible to group the interactions that were present on various databases, and therefore present a stronger evidence than the first group for being ataxin-3 interactors. Additionally, for those grouped interactors, PANTHER protein class enrichment analysis was deployed (**Table1**) revealing a dataset enriched in proteases and ubiquitin-binding ligases, which, according to the literature may indicate potential ataxin-3 interactors. From this work it was possible to select two groups of ataxin-3 interactors, those with strong support as ataxin-3 interactors and another group where evidence is not that strong for being ataxin-3 interactors, to be used in the *in-silico* analyses.

The *in-silico* analyses reported in Sousa e Silva et al. [106] have been performed using 3D structure predictions obtained from I-TASSER. Nevertheless, since 2021 for all proteins available at Uniprot database, there is a 3D structure prediction available at AlphaFold database [39]. The availability of such data allows to perform the *in-silico* methodology described in Rocha et al. [3] in a shorter time period (I-TASSER structure prediction takes about 6-8 hours for each protein). According to the TM score, Sousa e Silva et al. [106] report that 70% of the network studied show similar protein fold when comparing I-TASSER and AlphaFold inferred structures. Moreover, these authors show that the patterns observed when using I-TASSER structures are the same, irrespective of whether the TM-score of the alignment of I-TASSER and AlphaFold structures indicates a similar fold or not. In this work I address the effect of using different 3D

structures prediction in the *in-silico* results with the wt ataxin-3 form, as reported in Chapter 2. Both methodologies exhibited highly correlated results and very similar interaction patterns with overlapping interacting regions on the protein residues 47-57; 85-90; 99-110; 166-171; 175-191. According to the literature [106], ataxin-3 interactors tend to bind primarily in the ataxin-3 JD and this result was obtained using both I-TASSER and AlphaFold interactor structures. Nevertheless, in contrast to when using I-TASSER structures, when using AlphaFold structures, important discrepancies with what is reported in the literature are obtained. For instance, VCP (Gene ID 7415), and PCAF (Gene ID 8850) are reported to bind more the ataxin-3 C-terminal region than the JD [113], but the results obtained using AlphaFold structures suggest the opposite. Therefore, as performed in Chapter 3, Alphafold structure predictions should be used to perform a quick check on whether the interactor being considered is a true ataxin-3 interactor, but detailed analyses should be performed using I-TASSER structure predictions. Using this approach in Chapter 3, 83 novel ataxin-3 interactors were identified that bind mostly at the JD, as expected, since the majority of the ataxin-3 interactions occurs in this region [106]. These 83 proteins, show similar patterns to those observed in Chapter 2.

In this work I also address the effect of using two 3D structure predictions on binding strength differences when comparing wt and exp ataxin-3 interactions. The use of different 3D structure predictions when docking is performed with the exp ataxin-3 form show different results. Further work is needed to understand the causes of such differences. Therefore, for those ataxin-3 interactors that when using Alphafold 3D structure predictions show a larger number of interactions with the exp ataxin-3 form compared with the wt ataxin-3 form, the results should be confirmed using the I-TASSER structure prediction methodology. Using these criteria, here I identified 31 ataxin-3 interactors that show stronger interaction strength with the exp ataxin-3 form than with the wt ataxin-3 form. According to PANTHER GO-Slim Cellular Component enrichment analysis these proteins are significantly enrichment in mitochondria membrane proteins [117], proteins related to ubiquitin-protease systems [4, 119] and proteins located in the cytoplasm [26, 106], which according to the literature are components characteristic of ataxin-3 interactors, or present in the SCA3 pathology. These proteins should be studied in detail, using for instance *D. melanogaster* as a model species, in order to validate these inferences.

According to the work of Rocha, et al. [3] the presence of six or more interaction in the polyQ stretch in ATXN-1, is associated with wrong binding predictions. Therefore, Sousa

e Silva, et al. [106] discarded those proteins that present such pattern. Here I show that the presence of six or more interactions in the polyQ region do not necessarily imply a disagreement between the I-TASSER and the AlphaFold approaches and thus, can be included in the analysis.

Both Ataxin-3 I-TASSER and AlphaFold 3D structures predictions are supported by experimental evidence. Here I show that the interaction pattern observed is different with the two ataxin-3 structures for 80 % of the ataxin-3 interactors studied. Since when using AlphaFold 3D structure only 20 % of the ataxin-3 interactors show binding at the JD, a region where most of the ataxin-3 interactors bind according to the literature (see references in Sousa e Silva et al. [106]), and that is observed when using I-TASSER ataxin-3 structure prediction, in future work the latter prediction should be used.

Bibliographic references

1. McLoughlin, H.S., L.R. Moore, and H.L. Paulson, *Pathogenesis of SCA3 and implications for other polyglutamine diseases*. Neurobiology of disease, 2020. **134**: p. 104635.
2. Lim, J., et al., *A protein–protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration*. Cell, 2006. **125**(4): p. 801-814.
3. Rocha, S., et al., *ATXN1 N-terminal region explains the binding differences of wild-type and expanded forms*. BMC medical genomics, 2019. **12**(1): p. 1-14.
4. Liu, H., et al., *The Machado–Joseph disease deubiquitinase ataxin-3 regulates the stability and apoptotic function of p53*. PLoS biology, 2016. **14**(11): p. e2000733.
5. Zeng, L., et al., *Loss of the Spinocerebellar Ataxia type 3 disease protein ATXN3 alters transcription of multiple signal transduction pathways*. PLoS One, 2018. **13**(9): p. e0204438.
6. Vázquez, N., et al. *EvoPPI: A Web Application to Compare Protein-Protein Interactions (PPIs) from Different Databases and Species*. in *Practical Applications of Computational Biology and Bioinformatics, 12th International Conference*. 2019. Springer.
7. Cummings, C.J. and H.Y. Zoghbi, *Fourteen and counting: unraveling trinucleotide repeat diseases*. Human molecular genetics, 2000. **9**(6): p. 909-916.
8. Gkekas, I., et al., *Oxidative stress and neurodegeneration: Interconnected processes in polyq diseases*. Antioxidants, 2021. **10**(9): p. 1450.
9. Santana, M.M., et al., *A standardized protocol for blood and cerebrospinal fluid collection and processing for biomarker research in ataxia*. Neuropathology and Applied Neurobiology, 2023: p. e12892.
10. Chakraborty, A., et al., *Deficiency in classical nonhomologous end-joining–mediated repair of transcribed genes is linked to SCA3 pathogenesis*. Proceedings of the National Academy of Sciences, 2020. **117**(14): p. 8154-8165.
11. Li, Q.-F., et al., *Clinical features and genetic characteristics of homozygous spinocerebellar ataxia type 3*. Molecular Genetics & Genomic Medicine, 2020. **8**(9): p. e1314.
12. Iannuzzelli, K., et al., *The association between educational attainment and SCA 3 age of onset and disease course*. Parkinsonism & Related Disorders, 2022. **98**: p. 99-102.
13. Li, J., et al., *DNA methylation age acceleration is associated with age of onset in Chinese spinocerebellar ataxia type 3 patients*. Neurobiology of Aging, 2022. **113**: p. 1-6.
14. Mendonça, N., et al., *Clinical features of Machado-Joseph disease*. Polyglutamine Disorders, 2018: p. 255-273.
15. Riess, O., et al., *SCA3: neurological features, pathogenesis and animal models*. The Cerebellum, 2008. **7**: p. 125-137.
16. Matos, C.A., L.P. de Almeida, and C. Nóbrega, *Machado–Joseph disease/spinocerebellar ataxia type 3: lessons from disease pathogenesis and clues into therapy*. Journal of neurochemistry, 2019. **148**(1): p. 8-28.

17. Mao, Y., et al., *Deubiquitinating function of ataxin-3: insights from the solution structure of the Josephin domain*. Proceedings of the National Academy of Sciences, 2005. **102**(36): p. 12700-12705.
18. Sicorello, A., et al., *Capturing the conformational ensemble of the mixed folded polyglutamine protein ataxin-3*. Structure, 2021. **29**(1): p. 70-81. e5.
19. Weishäupl, D., et al., *Physiological and pathophysiological characteristics of ataxin-3 isoforms*. Journal of Biological Chemistry, 2019. **294**(2): p. 644-661.
20. Harris, G.M., et al., *Splice isoforms of the polyglutamine disease protein ataxin-3 exhibit similar enzymatic yet different aggregation properties*. PloS one, 2010. **5**(10): p. e13695.
21. Bettencourt, C., et al., *Increased transcript diversity: novel splicing variants of Machado–Joseph Disease gene (ATXN3)*. Neurogenetics, 2010. **11**: p. 193-202.
22. Schuster, K.H., et al., *Impaired oligodendrocyte maturation is an early feature in SCA3 disease pathogenesis*. Journal of Neuroscience, 2022. **42**(8): p. 1604-1617.
23. Da Silva, J.D., A. Teixeira-Castro, and P. Maciel, *From pathogenesis to novel therapeutics for spinocerebellar ataxia type 3: evading potholes on the way to translation*. Neurotherapeutics, 2019. **16**: p. 1009-1031.
24. Schuster, K.H., A.F. Putka, and H.S. McLoughlin, *Pathogenetic Mechanisms Underlying Spinocerebellar Ataxia Type 3 Are Altered in Primary Oligodendrocyte Culture*. Cells, 2022. **11**(16): p. 2615.
25. Temussi, P.A., G.G. Tartaglia, and A. Pastore, *The seesaw between normal function and protein aggregation: How functional interactions may increase protein solubility*. Bioessays, 2021. **43**(6): p. 2100031.
26. do Carmo Costa, M. and H.L. Paulson, *Toward understanding Machado–Joseph disease*. Progress in neurobiology, 2012. **97**(2): p. 239-257.
27. Matos, C.A., S. de Macedo-Ribeiro, and A.L. Carvalho, *Polyglutamine diseases: the special case of ataxin-3 and Machado–Joseph disease*. Progress in neurobiology, 2011. **95**(1): p. 26-48.
28. Bevivino, A.E. and P.J. Loll, *An expanded glutamine repeat destabilizes native ataxin-3 structure and mediates formation of parallel β -fibrils*. Proceedings of the National Academy of Sciences, 2001. **98**(21): p. 11955-11960.
29. Figueiredo, F., et al., *A Robust Assay to Monitor Ataxin-3 Amyloid Fibril Assembly*. Cells, 2022. **11**(12): p. 1969.
30. Fricano, A. and M. Midiri, *UNDERSTANDING THE RELATIONSHIP BETWEEN NORMAL FUNCTION AND ABERRANT AGGREGATION: THE CASE OF ATAXIN-3*. 2020.
31. Winborn, B.J., et al., *The deubiquitinating enzyme ataxin-3, a polyglutamine disease protein, edits Lys63 linkages in mixed linkage ubiquitin chains*. Journal of Biological Chemistry, 2008. **283**(39): p. 26436-26443.
32. Almeida, B., et al., *Trinucleotide repeats: a structural perspective*. Frontiers in neurology, 2013. **4**: p. 76.
33. Sawal, H.A., et al., *Comparative In Silico Analysis and Functional Characterization of TANK-Binding Kinase 1–Binding Protein 1*. Bioinformatics and Biology Insights, 2023. **17**: p. 11779322231164828.
34. Information, B.M.N.L.o.M.U.N.C.f.B. *National Center for Biotechnology Information*. 2004 [cited 2023 10/04]; Available from: <https://www.ncbi.nlm.nih.gov/gene/>.
35. Kuhlman, B. and P. Bradley, *Advances in protein structure prediction and design*. Nature Reviews Molecular Cell Biology, 2019. **20**(11): p. 681-697.
36. Chen, C.-C., J.-K. Hwang, and J.-M. Yang, *2-v2: template-based protein structure prediction server*. BMC Bioinformatics, 2009. **10**: p. 1-13.

37. Dhingra, S., et al., *A glance into the evolution of template-free protein structure prediction methodologies*. Biochimie, 2020. **175**: p. 85-92.
38. Zheng, W., et al., *Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations*. Cell reports methods, 2021. **1**(3): p. 100014.
39. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold*. Nature, 2021. **596**(7873): p. 583-589.
40. Kryshchak, A., et al., *Critical assessment of methods of protein structure prediction (CASP)—Round XIII*. Proteins: Structure, Function, and Bioinformatics, 2019. **87**(12): p. 1011-1020.
41. Ruff, K.M. and R.V. Pappu, *AlphaFold and implications for intrinsically disordered proteins*. Journal of Molecular Biology, 2021. **433**(20): p. 167208.
42. Kryshchak, A., et al., *Critical assessment of methods of protein structure prediction (CASP)—Round XIV*. Proteins: Structure, Function, and Bioinformatics, 2021. **89**(12): p. 1607-1617.
43. David, A., et al., *The AlphaFold database of protein structures: a biologist's guide*. Journal of molecular biology, 2022. **434**(2): p. 167336.
44. Wei, G.-W., *Protein structure prediction beyond AlphaFold*. Nature Machine Intelligence, 2019. **1**(8): p. 336-337.
45. Saxe, A., S. Nelli, and C. Summerfield, *If deep learning is the answer, what is the question?* Nature Reviews Neuroscience, 2021. **22**(1): p. 55-67.
46. Postic, G., et al., *MyPMFs: a simple tool for creating statistical potentials to assess protein structural models*. Biochimie, 2018. **151**: p. 37-41.
47. Varadi, M., et al., *AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models*. Nucleic acids research, 2022. **50**(D1): p. D439-D444.
48. Wu, S., J. Skolnick, and Y. Zhang, *Ab initio modeling of small proteins by iterative TASSER simulations*. BMC biology, 2007. **5**: p. 1-10.
49. Zhang, Y. and J. Skolnick, *Automated structure prediction of weakly homologous proteins on a genomic scale*. Proceedings of the National Academy of Sciences, 2004. **101**(20): p. 7594-7599.
50. Zhou, X., et al., *I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction*. Nature Protocols, 2022. **17**(10): p. 2326-2353.
51. Zhang, Y., *I-TASSER server for protein 3D structure prediction*. BMC bioinformatics, 2008. **9**: p. 1-8.
52. Wang, S., et al., *Protein-protein interaction networks as miners of biological discovery*. Proteomics, 2022. **22**(15-16): p. 2100190.
53. Chatr-Aryamontri, A., et al., *The BioGRID interaction database: 2017 update*. Nucleic acids research, 2017. **45**(D1): p. D369-D379.
54. Licata, L., et al., *MINT, the molecular interaction database: 2012 update*. Nucleic acids research, 2012. **40**(D1): p. D857-D861.
55. Alanis-Lobato, G., M.A. Andrade-Navarro, and M.H. Schaefer, *HIPPIE v2. 0: enhancing meaningfulness and reliability of protein-protein interaction networks*. Nucleic acids research, 2016: p. gkw985.
56. Sousa, A., et al., *On the identification of potential novel therapeutic targets for spinocerebellar ataxia type 1 (SCA1) neurodegenerative disease using EvoPPI3*. Journal of Integrative Bioinformatics, 2023(0): p. 20220056.
57. Sun, M.G. and P.M. Kim, *Evolution of biological interaction networks: from models to real data*. Genome biology, 2011. **12**(12): p. 1-11.
58. Vázquez, N., et al., *EvoPPI 1.0: a web platform for within-and between-species multiple Interactome comparisons and application to nine PolyQ proteins determining neurodegenerative diseases*. Interdisciplinary Sciences: Computational Life Sciences, 2019. **11**: p. 45-56.

59. Reboiro-Jato, M., et al. *EvoPPI 2: A Web and Local Platform for the Comparison of Protein–Protein Interaction Data from Multiple Sources from the Same and Distinct Species*. in *Practical Applications of Computational Biology and Bioinformatics, 16th International Conference (PACBB 2022)*. 2022. Springer.
60. Mi, H., et al., *PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools*. *Nucleic acids research*, 2019. **47**(D1): p. D419-D426.
61. Fisher, R.A., *On the interpretation of χ^2 from contingency tables, and the calculation of P*. *Journal of the royal statistical society*, 1922. **85**(1): p. 87-94.
62. Bonferroni, C., *Teoria statistica delle classi e calcolo delle probabilita*. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 1936. **8**: p. 3-62.
63. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing*. *Journal of the Royal statistical society: series B (Methodological)*, 1995. **57**(1): p. 289-300.
64. Jones, S. and J.M. Thornton, *Principles of protein-protein interactions*. *Proceedings of the National Academy of Sciences*, 1996. **93**(1): p. 13-20.
65. Rao, V.S., et al., *Protein-protein interaction detection: methods and analysis*. *International journal of proteomics*, 2014. **2014**.
66. Richards, A.L., M. Eckhardt, and N.J. Krogan, *Mass spectrometry-based protein–protein interaction networks for the study of human diseases*. *Molecular systems biology*, 2021. **17**(1): p. e8792.
67. Kristensen, L.V., et al., *Mass spectrometry analyses of normal and polyglutamine expanded ataxin-3 reveal novel interaction partners involved in mitochondrial function*. *Neurochemistry International*, 2018. **112**: p. 5-17.
68. Piasecki, P., et al., *Impaired interactions of ataxin-3 with protein complexes reveals their specific structure and functions in SCA3 Ki150 model*. *Frontiers in Molecular Neuroscience*, 2023.
69. Abeditashi, M., et al., *KPNB1 modulates the Machado–Joseph disease protein ataxin-3 through activation of the mitochondrial protease CLPP*. *Cellular and Molecular Life Sciences*, 2022. **79**(8): p. 401.
70. Sunny, S. and P. Jayaraj, *Protein–protein docking: Past, present, and future*. *The protein journal*, 2022: p. 1-26.
71. Katchalski-Katzir, E., et al., *Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques*. *Proceedings of the National Academy of Sciences*, 1992. **89**(6): p. 2195-2199.
72. Camacho, C.J., et al., *Scoring docked conformations generated by rigid-body protein-protein docking*. *Proteins: Structure, Function, and Bioinformatics*, 2000. **40**(3): p. 525-537.
73. Desta, I.T., et al., *Performance and its limits in rigid body protein-protein docking*. *Structure*, 2020. **28**(9): p. 1071-1081. e3.
74. Vajda, S. and D. Kozakov, *Convergence and combination of methods in protein–protein docking*. *Current opinion in structural biology*, 2009. **19**(2): p. 164-170.
75. Koshland Jr, D., *Application of a theory of enzyme specificity to protein synthesis*. *Proceedings of the National Academy of Sciences*, 1958. **44**(2): p. 98-104.
76. Bonvin, A.M., *Flexible protein–protein docking*. *Current opinion in structural biology*, 2006. **16**(2): p. 194-200.
77. Andrusier, N., et al., *Principles of flexible protein–protein docking*. *Proteins: Structure, Function, and Bioinformatics*, 2008. **73**(2): p. 271-289.

78. Van Zundert, G., et al., *The HADDOCK2. 2 web server: user-friendly integrative modeling of biomolecular complexes*. Journal of molecular biology, 2016. **428**(4): p. 720-725.
79. Honorato, R.V., et al., *Structural biology in the clouds: the WeNMR-EOSC ecosystem*. Frontiers in Molecular Biosciences, 2021. **8**: p. 729513.
80. Dominguez, C., R. Boelens, and A.M. Bonvin, *HADDOCK: a protein–protein docking approach based on biochemical or biophysical information*. Journal of the American Chemical Society, 2003. **125**(7): p. 1731-1737.
81. van Dijk, M., et al., *Information-driven protein–DNA docking using HADDOCK: it is a matter of flexibility*. Nucleic acids research, 2006. **34**(11): p. 3317-3325.
82. de Vries, S.J. and A.M. Bonvin, *CPORT: a consensus interface predictor and its performance in prediction-driven docking with HADDOCK*. PloS one, 2011. **6**(3): p. e17695.
83. Koukos, P.I., et al., *An overview of data-driven HADDOCK strategies in CAPRI rounds 38-45*. Proteins: Structure, Function, and Bioinformatics, 2020. **88**(8): p. 1029-1036.
84. Oughtred, R., et al., *The BioGRID interaction database: 2019 update*. Nucleic acids research, 2019. **47**(D1): p. D529-D541.
85. Stark, C., et al., *BioGRID: a general repository for interaction datasets*. Nucleic acids research, 2006. **34**(suppl_1): p. D535-D539.
86. Luck, K., et al., *A reference map of the human binary protein interactome*. Nature, 2020. **580**(7803): p. 402-408.
87. Murali, T., et al., *DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila*. Nucleic acids research, 2011. **39**(suppl_1): p. D736-D743.
88. Attrill, H., et al., *FlyBase: establishing a Gene Group resource for Drosophila melanogaster*. Nucleic acids research, 2016. **44**(D1): p. D786-D792.
89. López, Y., K. Nakai, and A. Patil, *HitPredict version 4: comprehensive reliability scoring of physical protein–protein interactions from more than 100 species*. Database, 2015. **2015**.
90. Persico, M., et al., *HomoMINT: an inferred human network based on orthology mapping of protein interactions discovered in model organisms*. BMC bioinformatics, 2005. **6**(4): p. 1-12.
91. Meyer, M.J., et al., *INstruct: a database of high-quality 3D structurally resolved protein interactome networks*. Bioinformatics, 2013. **29**(12): p. 1577-1579.
92. Mosca, R., A. Céol, and P. Aloy, *Interactome3D: adding structural details to protein networks*. Nature methods, 2013. **10**(1): p. 47-53.
93. Calderone, A., L. Castagnoli, and G. Cesareni, *Mentha: a resource for browsing integrated protein-interaction networks*. Nature methods, 2013. **10**(8): p. 690-691.
94. Zanzoni, A., et al., *MINT: a Molecular INTERaction database*. FEBS letters, 2002. **513**(1): p. 135-140.
95. Cowley, M.J., et al., *PINA v2. 0: mining interactome modules*. Nucleic acids research, 2012. **40**(D1): p. D862-D865.
96. Huang, H., et al., *Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes*. Genome biology, 2004. **5**(7): p. 1-15.
97. Schaefer, M.H., E.E. Wanker, and M.A. Andrade-Navarro, *Evolution and function of CAG/polyglutamine repeats in protein–protein interaction networks*. Nucleic acids research, 2012. **40**(10): p. 4273-4287.
98. La Spada, A.R. and J.P. Taylor, *Repeat expansion disease: progress and puzzles in disease pathogenesis*. Nature Reviews Genetics, 2010. **11**(4): p. 247-258.

99. Petrakis, S., et al., *Aggregation of polyQ-extended proteins is promoted by interaction with their natural coiled-coil partners*. Bioessays, 2013. **35**(6): p. 503-507.
100. Totzeck, F., M.A. Andrade-Navarro, and P. Mier, *The protein structure context of polyQ regions*. PloS one, 2017. **12**(1): p. e0170801.
101. Chavali, S., et al., *Amino acid homorepeats in proteins*. Nature Reviews Chemistry, 2020. **4**(8): p. 420-434.
102. Mier, P. and M.A. Andrade-Navarro, *Between interactions and aggregates: the PolyQ balance*. Genome biology and evolution, 2021. **13**(11): p. evab246.
103. Seidel, K., et al., *Brain pathology of spinocerebellar ataxias*. Acta neuropathologica, 2012. **124**: p. 1-21.
104. Uhlen, M. *A human protein atlas*. in *The FEBS Journal*. 2009. Wiley.
105. IBM, C., *Released 2020*. IBM SPSS Statistics for Windows, Version 27.0.
106. Sousa e Silva, R., et al., *The Josephin domain (JD) containing proteins are predicted to bind to the same interactors: Implications for spinocerebellar ataxia type 3 (SCA3) studies using Drosophila melanogaster mutants*. Frontiers in Molecular Neuroscience, 2023. **16**: p. 1140719.
107. Haacke, A., F.U. Hartl, and P. Breuer, *Calpain inhibition is sufficient to suppress aggregation of polyglutamine-expanded ataxin-3*. Journal of Biological Chemistry, 2007. **282**(26): p. 18851-18856.
108. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold*. Nature, 2021. **596**(7873): p. 583-589.
109. Roy, A., A. Kucukural, and Y. Zhang, *I-TASSER: a unified platform for automated protein structure and function prediction*. Nat Protoc, 2010. **5**(4): p. 725-38.
110. Yang, J. and Y. Zhang, *I-TASSER server: new development for protein structure and function predictions*. Nucleic Acids Res, 2015. **43**(W1): p. W174-81.
111. Zhang, Y. and J. Skolnick, *TM-align: a protein structure alignment algorithm based on the TM-score*. Nucleic acids research, 2005. **33**(7): p. 2302-2309.
112. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state*. Journal of molecular biology, 2007. **372**(3): p. 774-797.
113. Matos, C.A., L.P.d. Almeida, and C. Nóbrega, *Machado-Joseph disease/spinocerebellar ataxia type 3: lessons from disease pathogenesis and clues into therapy*. Journal of Neurochemistry, 2019. **148**(1): p. 8-28.
114. Petrakis, S., et al., *Aggregation of polyQ-extended proteins is promoted by interaction with their natural coiled-coil partners*. Bioessays, 2013. **35**(6): p. 503-7.
115. Schaefer, M.H., E.E. Wanker, and M.A. Andrade-Navarro, *Evolution and function of CAG/polyglutamine repeats in protein-protein interaction networks*. Nucleic Acids Res, 2012. **40**(10): p. 4273-87.
116. Mier, P. and M.A. Andrade-Navarro, *Between Interactions and Aggregates: The PolyQ Balance*. Genome Biol Evol, 2021. **13**(11).
117. Chou, A.-H., et al., *Polyglutamine-expanded ataxin-3 activates mitochondrial apoptotic pathway by upregulating Bax and downregulating Bcl-xL*. Neurobiology of disease, 2006. **21**(2): p. 333-345.
118. Neves-Carvalho, A., et al., *Regulation of neuronal mRNA splicing and Tau isoform ratio by ATXN3 through deubiquitylation of splicing factors*. BioRxiv, 2019: p. 711424.
119. Jana, N.R., et al., *Co-chaperone CHIP associates with expanded polyglutamine protein and promotes their degradation by proteasomes*. Journal of Biological Chemistry, 2005. **280**(12): p. 11635-11640.
120. Rizo, J. and J. Xu, *The synaptic vesicle release machinery*. Annual review of biophysics, 2015. **44**: p. 339-367.

121. Jeans, A.F., et al., *A dominant mutation in Snap25 causes impaired vesicle trafficking, sensorimotor gating, and ataxia in the blind-drunk mouse*. Proceedings of the National Academy of Sciences, 2007. **104**(7): p. 2431-2436.
122. Shen, X.-M., et al., *Mutant SNAP25B causes myasthenia, cortical hyperexcitability, ataxia, and intellectual disability*. Neurology, 2014. **83**(24): p. 2247-2255.
123. Fontaine, S.N., et al., *DnaJ/Hsc70 chaperone complexes control the extracellular release of neurodegenerative-associated proteins*. The EMBO journal, 2016. **35**(14): p. 1537-1549.
124. Chan, W.M., et al., *Expanded polyglutamine domain possesses nuclear export activity which modulates subcellular localization and toxicity of polyQ disease protein via exportin-1*. Human molecular genetics, 2011. **20**(9): p. 1738-1750.
125. Piasecki, P., et al., *Impaired interactions of ataxin-3 with protein complexes reveals their specific structure and functions in SCA3 Ki150 model*. Frontiers in Molecular Neuroscience, 2023. **16**: p. 1122308.
126. Nicastro, G., et al., *The solution structure of the Josephin domain of ataxin-3: structural determinants for molecular recognition*. Proceedings of the National Academy of Sciences, 2005. **102**(30): p. 10493-10498.
127. Wang, Y. and G.A. Voth, *Molecular dynamics simulations of polyglutamine aggregation using solvent-free multiscale coarse-grained models*. The journal of physical chemistry B, 2010. **114**(26): p. 8735-8743.
128. Satoh, T., et al., *Mode of substrate recognition by the Josephin domain of ataxin-3, which has an endo-type deubiquitinase activity*. FEBS letters, 2014. **588**(23): p. 4422-4430.
129. Song, A.-X., et al., *Structural transformation of the tandem ubiquitin-interacting motifs in ataxin-3 and their cooperative interactions with ubiquitin chains*. PLoS One, 2010. **5**(10): p. e13202.
130. Wen, J., D.R. Scoles, and J.C. Facelli, *Effects of the enlargement of polyglutamine segments on the structure and folding of ataxin-2 and ataxin-3 proteins*. Journal of Biomolecular Structure and Dynamics, 2017. **35**(3): p. 504-519.
131. Berman, H.M., et al., *The protein data bank*. Nucleic acids research, 2000. **28**(1): p. 235-242.
132. G. Nicastro, L.M., V. Esposito, R. Menon, A. Pastore. *Solution structure of the Josephin domain of Ataxin-3 in complex with ubiquitin molecule*. 2008 28/09/2023

]; Available from: <https://www.rcsb.org/structure/2JRI>.

Appendix

Supplementary Table 1. Representation of Ataxin-3 interactors in the main databases

Presence in main	Ataxin-3 interactors GeneID (gene name)
1	836 (CASP3); 3304 (CCDS34415.1); 7335 (CCDS13426.1); 9314 (CCDS6770.2); 11337 (ATG8A); 6128 (CCDS9162.1); 9927 (CCDS30587.1); 55742 (CCDS44541.2); 9894 (CCDS32363.1); 8022 (CCDS6994.1); 440738 (ATG8J); 387522 (CCDS13424.1)
2	87 (ACTN1); 834 (CASP1); 3303 (CCDS34414.1); 7334 (CCDS31875.1); 5727 (BCNS); 10422 (CCDS35177.1); 84557 (ATG8E); 7326 (CCDS32532.1)
3	31 (ACAC); 81 (ACTININ-4); 142 (ADPRT); 191 (adoHcyase); 226 (ALDA); 329 (API1); 372 (ARCN1); 439 (ARSA-I); 476 (ATP1A1); 498 (ATP5A); 506 (ATP5B); 509 (ATP5C); 515 (ATP5F1); 539 (ATP5O); 705 (BYSL); 708 (C1QBP); 790 (CAD); 840 (CASP-7); 908 (CCDS34640.1); 1072 (CCDS8114.1); 1152 (B-CK); 1213 (CCDS32696.1); 1314 (AILJK); 1434 (CAS); 1537 (CCDS6415.1); 1615 (aspRS); 1629 (BCATE2); 1656 (CCDS44751.1); 1738 (CCDS5749.1); 1743 (CCDS9833.1); 1828 (CCDS11896.1); 1832 (CCDS4501.1); 1933 (CCDS2367.1); 1936 (CCDS47930.1); 1937 (CCDS44626.1); 1938 (CCDS12117.1); 1981 (CCDS3259.1); 2023 (CCDS97.1); 2058 (CCDS31027.1); 2194 (CCDS11801.1); 2316 (ABP-280); 2597 (CCDS58201.1); 2618 (AIRS); 2969 (BAP135); 3029 (CCDS10447.2); 3064 (CCDS43206.1); 3181 (CCDS43557.1); 3190 (AUKS); 3191 (CCDS33015.1); 3192 (CCDS31081.1); 3276 (ANM1); 3305 (CCDS34413.1); 3309 (BIP); 3310 (CCDS1231.1); 3313 (CCDS4208.1); 3315 (CCDS5583.1); 3326 (CCDS4909.1); 3329 (CCDS33357.1); 3336 (CCDS2320.1); 3376 (CCDS6694.1); 3615 (CCDS2786.1); 3735 (CCDS10923.1); 3837 (CCDS11513.1); 3849 (CCDS8835.1); 3910 (CCDS34514.1); 3920 (CCDS14599.1); 3921 (CCDS2686.1); 3945 (CCDS8691.1); 3980 (CCDS11284.2); 4085 (CCDS3715.1); 4141 (CCDS8942.1); 4282 (CCDS13819.1); 10935 (AOP-1); 7203 (CCDS1140.2); 6510 (AAAT); 10540 (CCDS44930.1); 10606 (ADE2); 5717 (CCDS11272.1); 9747 (CCDS56514.1); 4627 (BDPLT6); 23191 (CCDS73695.1); 124801 (CCDS11475.1); 6279 (CAGA); 10342 (CCDS2939.1); 80273 (CCDS3396.1); 10574 (CCDS42696.1); 6874 (CCDS33500.1); 5718 (CCDS11669.1); 5685 (CCDS10303.1); 6742 (CCDS5866.1); 4628 (CCDS11144.1); 7965 (AIMP2); 5917 (ArgRS); 5091 (CCDS8152.1); 10575 (CCDS33206.1); 9551 (ATP5J2); 5719 (CCDS44504.1); 5686 (CCDS55619.1); 23193 (CCDS41656.1); 51495 (B-IND1); 23451 (CCDS33356.1); 55658 (CCDS12039.1); 23450 (CCDS10894.1); 56616 (CCDS73541.1); 6208 (CCDS4307.1); 5052 (CCDS522.1); 7001 (CCDS12281.1); 7266 (CCDS45677.1); 7167 (CCDS53740.1); 6407 (CCDS13346.1); 7431 (CCDS7120.1); 7531 (CCDS11001.1); 29927 (CCDS3046.1); 55660 (CCDS46430.1); 6210 (CCDS10571.1); 10075 (ARF-BP1); 26227 (CCDS904.1); 5682 (CCDS31431.1); 11331 (BAP); 9349 (CCDS11330.1); 7277 (ALS22); 6187 (CCDS10452.1); 10382 (beta-5); 51144 (CCDS7905.1); 5692 (CCDS996.1); 51441 (CAHL); 51639 (CCDS1707.1); 10383 (Beta2); 55768 (CCDS33719.1); 5230 (CCDS14438.1); 6188 (CCDS58161.1); 10945 (CCDS12718.1); 5693 (CCDS45083.1); 8402 (CCDS11059.1); 6189 (CCDS3775.1); 10549 (AOE37-2); 7874 (CCDS32385.1); 10946 (CCDS428.1); 4670

Supplementary Table 1 (continuation). Representation of Ataxin-3 interactors in the main databases

	(CCDS12203.1); 5694 (CCDS11056.1); 4967 (AKGDH); 51608 (C7orf20); 6223 (CCDS12588.1); 5430 (hRPB220); 6421 (CCDS388.1); 10121 (ACTR1A); 4869 (B23); 6950 (CCDS43522.1); 10576 (CCDS55843.1); 5687 (CCDS61437.1); 5093 (CCDS1898.1); 6117 (CCDS11014.1); 348262 (CCDS45814.1); 10213 (CCDS46437.1); 84271 (CCDS14038.1); 8826 (CCDS10362.1); 8661 (CCDS7608.1); 4830 (AWD); 6217 (CCDS12535.1); 6184 (CCDS3051.1); 6118 (CCDS314.1); 7407 (CCDS34412.1); 5689 (CCDS34577.1); 23197 (CCDS34296.1); 8266 (CCDS14754.1); 10314 (CCDS2392.1); 6119 (CCDS5356.1); 51637 (C14orf166); 29968 (CCDS6659.1); 154796 (AMOT); 5690 (CCDS394.1); 5591 (CCDS75734.1); 6748 (CCDS14731.1); 11273 (A2D); 9093 (CCDS10515.1); 7186 (CCDS7013.1); 92609 (CCDS33023.1); 10456 (CCDS1064.1); 8607 (CCDS3047.1); 9532 (BAG-2); 347688 (bA631M21.2); 4841 (CCDS14410.1); 10985 (CCDS41847.1); 10787 (CCDS2287.1); 5700 (CCDS32139.1); 55215 (CCDS10349.2); 9500 (CCDS14337.1); 5701 (CCDS5731.1); 254170 (BMND12); 8773 (CCDS10087.1); 5702 (CCDS7935.1); 83889 (CCDS46063.1); 5901 (ARA24); 79892 (C10orf119); 10294 (CCDS10726.1); 7520 (CCDS2402.1); 6191 (CCDS14418.1); 6224 (CCDS55231.1); 9131 (AIF); 10155 (CCDS12985.1); 8471 (CCDS14544.1); 4673 (CCDS76581.1); 5928 (CCDS366.1); 55740 (CCDS31040.1); 7184 (CCDS9094.1); 23304 (bA49A4.1); 58517 (CCDS32113.1); 7514 (CCDS33205.1); 9529 (BAG-5); 6194 (CCDS6492.1); 7086 (CCDS2871.1); 9861 (CCDS2901.1); 5708 (CCDS3258.1); 8813 (CCDS13434.1); 10134 (BAP31); 10399 (CCDS34324.1); 6633 (CCDS33053.1); 79568 (C2orf47); 5709 (CCDS11356.1); 6138 (CCDS2640.1); 192111 (BXLBV68); 9343 (CCDS11489.1); 7295 (CCDS35103.1); 6634 (CCDS13828.1); 5478 (CCDS5494.1); 84790 (bcm948); 6206 (CCDS5164.1); 10963 (CCDS60827.1); 5216 (ALS18); 6207 (CCDS7823.1); 23020 (ASCC3L1); 5315 (CCDS32284.1); 10130 (CCDS1675.1); 5902 (CCDS13775.1); 7917 (BAG-6); 51520 (CCDS34265.1); 6232 (CCDS1059.1); 10527 (CCDS31425.1); 5704 (CCDS12547.1); 22948 (CCDS3877.1); 5110 (CCDS5219.2); 10594 (CCDS11010.1); 57532 (CCDS32600.1); 5970 (CCDS31609.1); 6301 (CCDS795.1); 5706 (CADP44); 10694 (C21orf112); 10992 (CCDS31612.1); 5707 (CCDS2482.1); 6169 (CCDS11696.1); 6202 (CCDS513.1); 5245 (CCDS11548.1); 6203 (CCDS12884.1)
4	328 (APE); 391 (ARHG); 823 (CANP); 871 (AsTP3); 1019 (CCDS8953.1); 1468 (CCDS11786.1); 3420 (CCDS13031.1); 6576 (CCDS13758.1); 51170 (CCDS3619.1); 10212 (BAT1); 6175 (CCDS9193.1); 84263 (C9orf99); 10999 (ACSVL4); 337867 (CCDS45064.1); 92667 (bA504H3.4); 23464 (CCDS13957.1); 4702 (CCDS6835.1); 414919 (C8orf82); 54708 (CCDS7420.1); 90550 (C10orf42); 81929 (CCDS32791.1); 10381 (beta-4); 8443 (CCDS1592.1); 112724 (CCDS42627.1); 23011 (CCDS9003.1); 65260 (C1orf163); 493856 (CCDS34040.1); 119559 (BCRM1); 51115 (CCDS34918.1); 6390 (CCDS176.1); 5631 (ARTS); 79888 (AGPAT10); 10059 (CCDS61095.1); 7317 (A1S9); 27146 (CCDS47033.1); 200205 (C1orf69); 8451 (CCDS41908.1); 10797 (CCDS1935.2); 56052 (ALG1); 23647 (ARFIP2); 10892 (CCDS11967.1); 6566 (CCDS858.1)
5	1111 (CCDS58191.1); 4214 (CCDS43318.1); 83737 (ADMFD); 9656 (CCDS34384.1); 23295 (CCDS42115.1); 57154 (CCDS34689.1); 57472 (CCDS4455.1); 8878 (A170)

Supplementary Table 1 (continuation). Representation of Ataxin-3 interactors in the main databases

6	41 (ACCN2); 47 (ACL); 267 (AMFR); 292 (AAC2); 293 (AAC3); 351 (AAA); 396 (ARHGDIA); 598 (Bcl-X); 1385 (CCDS2374.1); 1642 (CCDS31576.1); 1778 (CCDS9966.1); 2033 (CCDS14010.1); 3300 (CCDS2439.1); 3301 (CCDS6533.1); 3308 (APG-2); 3312 (CCDS44754.1); 10013 (CCDS14306.1); 8850 (CAF); 5250 (CCDS9065.1); 8454 (CCDS34772.1); 22824 (APG-1); 9611 (CCDS11175.1); 7332 (CCDS13790.1); 79751 (CCDS7715.1); 9524 (CCDS12313.1); 23597 (ACATE2); 4697 (CCDS5357.1); 10808 (CCDS66525.1); 4303 (AFX); 5371 (CCDS10255.1); 27338 (CCDS33114.1); 7189 (CCDS7901.1); 7314 (CCDS11177.1); 6389 (CCDS3853.1); 9531 (BAG-3); 84447 (CCDS31605.1); 7846 (B-ALPHA-1); 5710 (AF); 83932 (C1orf124); 7157 (BCC7); 8678 (ATG6)
7	824 (CANP2); 1460 (CCDS4712.1); 3320 (CCDS32160.1); 10277 (CCDS110.1); 6383 (CCDS6272.1); 7341 (CCDS2352.1)
8	310 (ANX7); 1026 (CAP20); 1387 (CBP); 1785 (CCDS32907.1); 2547 (CCDS14021.1); 3105 (Aw-33); 4792 (CCDS9656.1); 56159 (CCDS35323.1); 64219 (CCDS14392.1); 51773 (CCDS8253.1); 121536 (AEBP2); 6717 (CCDS47638.1); 91942 (B17.2L); 57159 (CCDS1745.2); 84675 (CCDS6184.1); 373863 (CCDS4236.1); 29979 (CCDS6663.1); 6195 (CCDS284.1); 11284 (AOA4); 203068 (CCDS4687.1); 8841 (CCDS4264.1); 57662 (CAMSAP3); 7083 (CCDS11754.1); 8569 (CCDS30705.1); 84676 (CCDS273.1); 387990 (CCDS9734.1); 27005 (CCDS30920.1); 84722 (CCDS30791.1); 54499 (CCDS1251.2); 117584 (CARP-2); 5705 (CCDS11645.1); 100287932 (CCDS73091.1); 9274 (BCL7C)
9	821 (CANX); 2130 (bK984G1.4); 78990 (C14orf137); 6520 (CCDS31588.1); 4738 (CCDS9621.1); 9463 (CCDS13965.1); 9001 (CCDS11406.1); 84986 (ARHGAP19); 83892 (BTBD28); 8975 (CCDS3235.1)
10	5713 (CCDS10910.1)
11	2932 (CCDS2996.1); 10273 (CCDS10419.1); 80262 (C16orf6)
13	7316 (CCDS9260.1); 4287 (ataxin-3))
14	5071 (AR-JP)
15	7415 (ALS14)
16	5887 (CCDS59138.1)
18	5886 (CCDS12289.1)

Supplementary Table 2. Ataxin-3 human interactors present in different EvoPPI3 predict interactomes databases

Interactome and Predict interactomes databases	GeneID
<i>Homo polyQ_22</i> interactome	31, 32, 47, 58, 59, 60, 70, 71, 72, 81, 87, 101, 142, 191, 226, 274, 292, 293, 302, 329, 372, 391, 439, 476, 488, 489, 498, 506, 509, 515, 539, 696, 705, 708, 765, 790, 821, 840, 871, 908, 1005, 1019, 1072, 1152, 1213, 1302, 1314, 1387, 1434, 1468, 1537, 1615, 1629, 1642, 1656, 1738, 1743, 1778, 1793, 1828, 1832, 1915, 1917, 1933, 1936, 1937, 1938, 1953, 1973, 1974, 1981, 2015, 2023, 2033, 2058, 2066, 2078, 2130, 2185, 2194, 2200, 2316, 2475, 2547, 2597, 2618, 2738, 2825, 2969, 3029, 3064, 3084, 3122, 3123, 3178, 3181, 3187, 3188, 3190, 3191, 3192, 3276, 3301, 3303, 3304, 3305, 3308, 3309, 3310, 3312, 3313, 3315, 3320, 3326, 3329, 3336, 3376, 3420, 3557, 3615, 3688, 3710, 3735, 3757, 3837, 3849, 3857, 3910, 3920, 3921, 3945, 3980, 4085, 4141, 4259, 4282, 4627, 4628, 4670, 4673, 4697, 4702, 4790, 4809, 4830, 4831, 4841, 4862, 4869, 4931, 4967, 5046, 5052, 5071, 5091, 5093, 5110, 5168, 5179, 5216, 5230, 5245, 5250, 5315, 5430, 5478, 5591, 5631, 5682, 5685, 5686, 5687, 5688, 5689, 5690, 5692, 5693, 5694, 5700, 5701, 5702, 5704, 5705, 5706, 5707, 5708, 5709, 5710, 5713, 5717, 5718, 5719, 5858, 5880, 5886, 5887, 5901, 5902, 5917, 5928, 5970, 6117, 6118, 6119, 6138, 6169, 6175, 6184, 6187, 6188, 6189, 6191, 6194, 6202, 6203, 6204, 6206, 6207, 6208, 6210, 6217, 6223, 6224, 6232, 6233, 6279, 6301, 6310, 6389, 6407, 6418, 6421, 6493, 6510, 6520, 6527, 6566, 6576, 6628, 6633, 6634, 6638, 6721, 6742, 6748, 6950, 7001, 7039, 7086, 7157, 7167, 7171, 7184, 7186, 7203, 7266, 7277, 7278, 7280, 7295, 7314, 7316, 7317, 7407, 7415, 7431, 7485, 7514, 7520, 7529, 7531, 7534, 7846, 7874, 7881, 7917, 7965, 8125, 8266, 8402, 8443, 8471, 8607, 8655, 8661, 8678, 8685, 8701, 8773, 8813, 8826, 8836, 8850, 8926, 9001, 9093, 9131, 9228, 9317, 9343, 9349, 9463, 9500, 9524, 9529, 9532, 9551, 9747, 9861, 9891, 9940, 10049, 10059, 10075, 10121, 10130, 10134, 10155, 10198, 10212, 10213, 10273, 10277, 10294, 10314, 10342, 10376, 10382, 10383, 10395, 10396, 10397, 10398, 10399, 10456, 10473, 10527, 10540, 10541, 10549, 10574, 10575, 10576, 10594, 10606, 10627, 10694, 10787, 10797, 10808, 10892, 10935, 10945, 10946, 10963, 10966, 10971, 10985, 10992, 10999, 11083, 11273, 11284, 11331, 11337, 22824, 22948, 23011, 23020, 23191, 23193, 23197, 23304, 23325, 23450, 23451, 23464, 23543, 23550, 23597, 23647, 26038, 26091, 26227, 26502, 27146, 27351, 29777, 29842, 29927, 29968, 50618, 51115, 51144, 51170, 51441, 51444, 51495, 51520, 51608, 51637, 51639, 53353, 54518, 54708, 54715, 54777, 54801, 55102, 55130, 55206, 55215, 55507, 55658, 55660, 55740, 55768, 55872, 56052, 56159, 56288, 56616, 56899, 57162, 57532, 57597, 57608, 58517, 59339, 64421, 65260, 66002, 79058, 79568, 79603, 79640, 79751, 79888, 79892, 79943, 80036, 80122, 80262, 80273, 81929, 83732, 83746, 83889, 84083, 84181, 84254, 84263, 84271, 84542, 84790, 84795, 84875, 84894, 84959, 84986, 90550, 91522, 92609, 92667, 103910, 112714, 112724, 112840, 113457, 115948, 117583, 119559, 124801, 127602, 139818, 140469, 140606, 140731, 143471, 144983, 146713, 148870, 149465, 150353, 150365, 154796, 157570, 192111, 196383, 196441, 200205, 203068, 219899, 221496, 221613, 254050, 255626, 256227, 260293, 284297, 284541, 285755, 337867, 339761, 344191, 347688,

Supplementary Table 2 (continuation). Ataxin-3 human interactors present in different EvoPPI3 predict interactomes

	347733, 348262, 387509, 401285, 414061, 414919, 440738, 493856, 646817, 654364
<i>Homo</i> modifiers interactome	1108, 6869, 26224, 55806, 2242, 5026, 9020, 59284, 4159, 5311, 25897, 5129, 51251, 10922, 23387
Predict interactome using <i>M. musculus</i> modifiers genes using DIOPT orthologies	8329, 23524, 381, 388536, 5224, 10575, 319101, 3891, 771, 3420, 10840, 55236, 113835, 72, 25902, 3712, 728378, 2770, 222183, 11170, 94081, 8330, 50810, 5093, 2819, 84303, 9651, 2949, 56953, 124540, 4134, 388336, 388468, 5354, 128240, 2785, 3305, 3045, 8525, 7534, 132204, 57282, 11168, 79744, 9054, 10243, 137886, 10376, 217, 50808, 5223, 386617, 23193, 85358, 10574, 100533105, 8724, 1843, 221656, 3013, 51138, 26232, 396, 114757, 728577, 83641, 10971, 526, 9914, 3371, 119, 509, 51606, 1289, 1809, 3889, 5096, 1159, 23066, 100129842, 10381, 257194, 3418, 9588, 6285, 9390, 9456, 2280, 204474, 5691, 55967, 338785, 2768, 3710, 70, 5955, 3028, 4832, 5757, 6154, 801, 116085, 2037, 51673, 2167, 10382, 2947, 6286, 7277, 8334, 1468, 100037417, 51276, 4833, 92235, 4701, 5824, 284612, 6815, 51540, 1743, 10247, 730291, 2653, 8331, 11238, 124404, 5226, 5160, 2783, 3303, 5094, 10313, 23196, 147923, 3612, 9454, 341359, 3043, 653857, 10776, 5689, 51207, 492, 29934, 8662, 112714, 8728, 55833, 4830, 475, 5161, 8332, 55567, 115290, 10314, 7143, 50944, 22933, 6416, 9455, 1434, 22999, 36, 83447, 2670, 3320, 90317, 5954, 55964, 4831, 55832, 8927, 55766, 1287, 1807, 1937, 11112, 10055, 7148, 284293, 5034, 9196, 148198, 653781, 3887, 23136, 10121, 8337, 25977, 338599, 5232, 6157, 730087, 7280, 27034, 6223, 6620, 51015, 3708, 4440, 56896, 55707, 80349, 377, 1645, 2035, 9131, 9990, 127124, 83849, 27299, 221662, 669, 8338, 7347, 29942, 85236, 1466, 4000, 55970, 80150, 4771, 8801, 4705, 8867, 10717, 5630, 51412, 91975, 2781, 6155, 55506, 440854, 27032, 10383, 51806, 7278, 8335, 85235, 5230, 57026, 490, 1400, 51079, 9987, 57687, 140735, 10120, 3009, 4967, 168417, 213, 3855, 81931, 7345, 8402, 51807, 10450, 8336, 1660, 1302, 1432, 522, 34, 6421, 55968, 100287399, 5694, 9922, 255220, 51350, 25981, 765, 26047, 8936, 5236, 225689, 7284, 7086, 5104, 3024, 7417, 121504, 8342, 285346, 375, 5501, 6624, 4841, 667, 404785, 8871, 90592, 10126, 152687, 11315, 488, 4128, 10391, 586, 6427, 23603, 8277, 8343, 91120, 9796, 10787, 8607, 9066, 2779, 3039, 10982, 27300, 245973, 829, 84909, 51678, 8339, 56171, 5499, 8604, 7415, 245972, 6622, 374928, 9793, 115827, 3853, 5962, 4905, 3007, 22941, 23007, 137362, 8340, 11313, 5500, 79029, 292, 8341, 162, 10587, 7416, 5698, 128312, 8803, 23205, 8671, 139741, 6623, 2778, 1738, 8775, 9898, 10096, 548596, 389, 48, 29916, 113026, 390502, 8048, 5274, 1267, 5604, 7652, 10493, 3006, 6860, 3850, 5803, 84617, 210, 5605, 9899, 22848, 5010, 54954, 10163, 10097, 5275, 9371, 79861, 11221, 1917, 7125, 291, 161, 8644, 57333, 6857, 8971, 762, 1282, 11151, 11085, 2793, 1152, 85012, 7386, 3021, 5272, 811, 9368, 2810, 1770, 4744, 8774, 654364, 794, 2940, 347688, 645974, 11086, 2631, 56934, 653505, 5273, 170506, 5339, 90957, 283871, 5537, 5603, 7453, 51121, 127829, 9699, 534, 23176, 826, 5608, 25824, 387, 84622, 83697, 55752, 23181, 7458, 1508, 1768, 2776, 3848, 51060, 125972, 549, 208, 3799, 2923, 1915, 8120, 1476, 50863, 3313, 5213, 151011, 4747, 85413, 29984, 10693, 9900, 105372280, 145264, 8050, 23708, 1280, 392490, 6861, 2791, 23576, 5276, 10165, 6135, 191, 1329, 346562, 2580, 57468, 760, 1832, 3149, 644591, 2271, 7589, 2938, 3068, 5607, 12, 873, 1459, 2288, 402665, 5211, 115207, 55617, 83959, 57731, 223, 8514, 22982, 377677, 7594, 25828, 5414,

Supplementary Table 2 (continuation). Ataxin-3 human interactors present in different EvoPPI3 predict interactomes

55888, 80332, 5480, 174, 2595, 515, 345651, 118980, 9114, 27347, 10105, 824, 143425, 93611, 22987, 54039, 5216, 2774, 6207, 7529, 5481, 129684, 1653, 2823, 3797, 4820, 6868, 7001, 54434, 89778, 317772, 125919, 726, 2173, 3181, 5217, 26027, 287, 6208, 80264, 54961, 5478, 374454, 3049, 3098, 220988, 23382, 3831, 84823, 89777, 133121, 4355, 59, 57142, 5214, 6271, 9376, 22919, 84293, 57140, 64077, 5413, 1457, 5479, 51196, 2026, 2286, 3017, 55753, 3326, 400, 92815, 57471, 441234, 2936, 10434, 2806, 7532, 7334, 414149, 513, 65009, 80336, 54505, 1212, 8655, 8854, 11034, 388523, 27153, 3032, 10109, 201625, 23784, 57016, 4185, 2220, 57609, 64744, 7533, 805, 201161, 204, 2821, 5948, 5816, 3309, 1992, 10110, 55033, 2171, 7203, 81855, 723790, 10308, 9119, 728945, 1487, 7464, 9512, 10767, 2787, 3047, 9908, 5813, 199777, 51661, 51727, 7134, 339500, 7531, 92291, 10570, 80201, 10768, 1975, 3015, 1845, 51332, 26227, 398, 528, 5814, 127833, 22924, 219, 27020, 29993, 79937, 8362, 89953, 90019, 63923, 170487, 78986, 9551, 140625, 2786, 3046, 6446, 55335, 6512, 55269, 4729, 23028, 9948, 6711, 196385, 221357, 6050, 29767, 10146, 84932, 8363, 53616, 3728, 168374, 3014, 23228, 333932, 23096, 1844, 10808, 55268, 397, 6712, 3632, 6051, 348, 26528, 120, 10540, 1160, 8294, 9351, 8360, 10672, 3892, 127018, 6708, 8756, 6510, 64714, 770, 3419, 5519, 57447, 2281, 2639, 2769, 6709, 71, 25801, 284382, 4282, 11135, 29765, 83939, 51764, 8361, 5190, 10409, 646486, 84334, 163223, 9550, 7700, 55530, 1778, 57446, 2948, 6511, 105371242, 51303, 6710, 25934, 25802, 57116, 3306, 29766, 84929, 1290, 730262, 6252, 220296, 2784, 3304, 26270, 79849, 7111, 3613, 6318, 55207, 8366, 102724334, 6649, 493, 5526, 140886, 56925, 6583, 4931, 353088, 57718, 7178, 132430, 216, 476, 6187, 2752, 3856, 5064, 86, 84076, 3012, 8301, 6253, 8367, 866, 37, 6650, 525, 144983, 85464, 3191, 6584, 5527, 8632, 10944, 84671, 6782, 11076, 118, 378, 768, 1288, 1808, 8364, 347404, 100129543, 1158, 3890, 4731, 8761, 1499, 28976, 3417, 5590, 83481, 374887, 2767, 6647, 10016, 2036, 9356, 3336, 1646, 8365, 2946, 6317, 1109, 29968, 231, 5525, 6648, 7639, 10810, 101, 374291, 4133, 93661, 2782, 84940, 4340, 8370, 5265, 5331, 119559, 122664, 4082, 3042, 491, 5530, 6521, 5596, 30832, 7644, 6389, 6455, 10815, 5860, 23499, 8701, 646480, 440093, 23236, 3854, 7314, 5332, 441282, 25813, 9230, 3010, 90423, 6522, 8570, 35, 10486, 57060, 523, 5531, 5333, 7381, 30833, 6456, 23632, 147968, 4738, 3628, 5597, 3059, 643836, 766, 1286, 506, 6188, 3888, 8038, 162966, 8368, 5594, 445577, 3709, 10549, 5528, 6519, 10681, 285051, 4735, 6717, 81569, 170959, 2017, 4099, 203068, 126961, 9162, 18, 2944, 317754, 5330, 8369, 6189, 489, 1465, 9360, 126433, 5529, 587, 4001, 5595, 440689, 6520, 5793, 10814, 23564, 6718, 8766, 8969, 2780, 3040, 6921, 830, 50, 79659, 7846, 5864, 6855, 84812, 9498, 7318, 5336, 7384, 9630, 5600, 57528, 286436, 10753, 3008, 2098, 3852, 5865, 8970, 23108, 5931, 9829, 6856, 10093, 28988, 7385, 1431, 8508, 163, 5271, 293, 1301, 142679, 284323, 340543, 4478, 2194, 10949, 23171, 5862, 6853, 5928, 1284, 5267, 22907, 147700, 2665, 3837, 6457, 7514, 10487, 5532, 158931, 5334, 813, 5598, 9892, 1073, 3658, 4048, 6854, 79788, 8968, 6193, 56203, 487, 163227, 7184, 346288, 4129, 9497, 8506, 100528064, 7317, 5269, 162962, 5335, 445582, 5533, 10856, 763, 83660, 140807, 84717, 5042, 11252, 56904, 2664, 8346, 682, 116362, 160428, 1072, 27306, 9601, 3071, 5836, 2941, 4045, 10989, 6233, 8347, 89869, 64963, 94097, 51024, 146754, 7881, 10920, 84517, 51286, 83592, 388, 55845, 7419, 5239, 23406, 8344, 84253, 10590,

Supplementary Table 2 (continuation). Ataxin-3 human interactors present in different EvoPPI3 predict interactomes

	1769, 6560, 23208, 5834, 59345, 3851, 8939, 3005, 29089, 10128, 989, 653604, 100287226, 8345, 10327, 23473, 908, 5174, 92906, 4513, 9534, 160, 1737, 128853, 10723, 3184, 29887, 4716, 84457, 6632, 5575, 1281, 81616, 3101, 2792, 94235, 1622, 80824, 3020, 810, 9143, 761, 6500, 8350, 5245, 5708, 8747, 2939, 5510, 5576, 23546, 255626, 2029, 2289, 3329, 171392, 203102, 1850, 9144, 54331, 793, 147741, 9540, 2110, 84258, 8351, 10465, 22948, 5573, 2515, 554313, 55915, 2596, 4978, 100996331, 5837, 10131, 5903, 23410, 6300, 8348, 253512, 874, 825, 1767, 8546, 2645, 3849, 2775, 113457, 154150, 5375, 283106, 8745, 85445, 6697, 10793, 207, 808, 10000, 2824, 3800, 79763, 340252, 388743, 23411, 7094, 63908, 288, 1475, 10397, 8349, 3182, 3312, 94239, 4522, 5579, 653269, 10864, 3099, 2790, 91860, 347733, 3018, 221613, 759, 60, 8354, 3148, 5315, 57111, 158983, 6505, 7562, 286887, 7430, 645142, 2027, 2287, 653598, 8751, 112802, 84725, 5052, 553115, 91661, 9016, 284443, 11328, 8355, 102724488, 5250, 84064, 5577, 441531, 8748, 54928, 90141, 51228, 27445, 57835, 90075, 7757, 173, 55589, 8880, 9145, 11127, 84790, 2773, 3033, 5313, 388558, 823, 7295, 8352, 57175, 5578, 25794, 54927, 1652, 10863, 64837, 84656, 442721, 4719, 8749, 205, 3798, 51427, 10929, 5909, 3050, 8353, 1294, 286, 7296, 3310, 6508, 29763, 8358, 1488, 65108, 140679, 2788, 3048, 8754, 51102, 6574, 23157, 1277, 60682, 708, 284371, 10935, 4725, 3146, 58, 166378, 11331, 8359, 1846, 5518, 6509, 100506084, 5584, 57576, 6575, 60681, 6906, 529, 1537, 2805, 2935, 57180, 11332, 220, 6506, 5515, 382, 9545, 51100, 5713, 1211, 3421, 23154, 148213, 2023, 11131, 91664, 8290, 2771, 8356, 22824, 26986, 7167, 8224, 5516, 6507, 8357, 121391, 2950, 100526693, 203, 5582, 2820, 10801, 55526, 9943, 11132, 54205, 4137, 3178, 3308, 2170, 196374, 2739, 81027, 1731, 165324
Predict interactome using <i>M. musculus</i> modifiers genes using Ensembl orthologies	4729, 10840, 72, 9948, 84932, 8362, 23524, 6050, 9551, 10575, 2770, 5224, 25902, 29767, 53616, 6712, 348, 8363, 333932, 84303, 124540, 5093, 71, 396, 51138, 100506084, 57447, 11135, 11168, 9054, 29766, 2785, 3632, 5354, 128240, 9351, 8525, 2819, 10243, 132204, 5519, 26232, 4282, 3014, 10540, 8294, 8360, 57446, 23193, 10971, 83641, 6710, 8361, 217, 1160, 1778, 9550, 10574, 3046, 526, 5223, 10409, 8854, 23157, 10109, 513, 4725, 51764, 7532, 11331, 805, 2820, 60682, 3015, 58, 7334, 8358, 121504, 340543, 1277, 3421, 57609, 81855, 1992, 5816, 220, 5518, 11332, 6575, 708, 7203, 529, 1537, 8359, 51102, 29993, 7167, 203, 2805, 8224, 388336, 1487, 6506, 22824, 2821, 2771, 8356, 9545, 1211, 26986, 4137, 3032, 5814, 5516, 7531, 5582, 3178, 3308, 10768, 2739, 528, 8357, 1975, 26227, 54205, 80201, 26027, 5216, 6207, 2026, 2806, 515, 9114, 55888, 3017, 8354, 5315, 7001, 23154, 5052, 9908, 6208, 1653, 3033, 3098, 824, 5713, 84064, 7529, 5481, 5250, 2173, 2287, 8355, 11127, 5214, 23382, 205, 55589, 3831, 7295, 10929, 6271, 8352, 140735, 126961, 22919, 57175, 2773, 4355, 8880, 2823, 5313, 59, 5478, 27347, 3018, 1652, 3099, 286, 5909, 84293, 441531, 4719, 8353, 51228, 64837, 5413, 57471, 3050, 9143, 3182, 3312, 5245, 8350, 1622, 387, 100037417, 4716, 6500, 1508, 84457, 51326, 6632, 2110, 8351, 3799, 549, 255626, 25824, 5708, 54331, 1476, 1915, 5576, 10165, 3313, 23576, 12, 2271, 2596, 55915, 6135, 22948, 808, 873, 57468, 9900, 3020, 760, 1329, 1459, 7094, 6697, 8514, 2288, 7589, 63908, 8346, 115207, 79763, 56904, 3184, 10097, 5010, 389, 10493, 2810, 3005, 51024, 8775, 10856, 113457, 10096, 5604, 160428, 3021, 9898, 811, 8347, 1072, 10131, 8050, 1267, 1738, 7125, 291, 22848, 5803, 10989, 1917, 2923, 5605, 210, 7386, 7419, 55845,

Supplementary Table 2 (continuation). Ataxin-3 human interactors present in different EvoPPI3 predict interactomes

	6560, 2029, 794, 29089, 8344, 7881, 10920, 3071, 11151, 23406, 810, 85012, 10327, 5339, 2631, 7453, 160, 127829, 548596, 762, 5834, 8774, 1152, 23176, 51286, 10128, 4513, 4744, 7417, 586, 11315, 1431, 6624, 9630, 8342, 9498, 7384, 6921, 3007, 50, 7086, 28988, 7846, 7385, 4128, 6427, 8508, 488, 4478, 8607, 8343, 643836, 2098, 4841, 3039, 830, 9829, 10787, 201161, 7748, 10093, 8970, 3837, 7514, 10487, 5532, 8604, 2665, 11313, 6193, 7415, 3853, 10949, 2194, 9892, 9793, 5862, 682, 5928, 3024, 8506, 5500, 487, 127833, 8671, 5236, 7284, 7317, 162, 5269, 7416, 4048, 8968, 3040, 4905, 94239, 8803, 23205, 7184, 5596, 523, 3708, 6620, 10815, 7347, 80349, 6389, 1937, 3009, 11112, 6157, 4082, 5232, 377, 6223, 7280, 440093, 5860, 100996709, 25813, 56896, 10121, 55970, 10486, 30832, 5531, 80150, 5597, 90423, 8339, 7381, 6456, 3854, 2035, 9990, 23236, 286887, 8867, 27034, 25977, 4738, 4705, 9230, 9131, 30833, 5594, 57026, 522, 3709, 6717, 4099, 2017, 8402, 506, 6188, 6155, 5230, 57687, 8368, 133121, 1400, 1660, 5529, 4001, 5694, 489, 51079, 1465, 4340, 5265, 8370, 8766, 6520, 55506, 18, 6189, 9922, 23564, 440689, 213, 9162, 55968, 4967, 70, 525, 653604, 55207, 353088, 140886, 3320, 509, 801, 3336, 6252, 220296, 6285, 8366, 9456, 5955, 56925, 654364, 5096, 7111, 55967, 37, 492, 3417, 3612, 147923, 4701, 5824, 10944, 2752, 1159, 7178, 3628, 2037, 6187, 11076, 8367, 51606, 5064, 1434, 1499, 5689, 8662, 29968, 4731, 4830, 23196, 3191, 2783, 6647, 8761, 94081, 118, 378, 5226, 1743, 10247, 2653, 1808, 8364, 2947, 9454, 50944, 554313, 2767, 5160, 29934, 28976, 83447, 231, 3028, 22933, 4831, 22999, 8927, 6648, 10810, 1158, 2036, 475, 8365, 55832, 374291, 55964, 2670, 7143
Predict interactome using <i>D. rerio</i> predict polyQ_22 database using DIOPT orthologies	3047, 5222, 1487, 57282, 5354, 27288, 1666, 2868, 64284, 124404, 2949, 5223, 7073, 100526693, 89, 5091, 10376, 23191, 3030, 366, 2770, 22795, 3908, 7072, 204474, 3371, 3891, 54509, 5224, 6017, 55170, 1244, 4070, 8659, 161003, 3192, 114757, 50809, 6809, 171177, 51602, 481, 5093, 5357, 5225, 302, 822, 3712, 10642, 204, 2220, 50940, 9717, 404785, 8924, 50808, 6810, 79612, 56886, 59, 5094, 101060233, 3695, 132204, 440854, 220988, 5358, 55304, 341359, 2318, 1278, 400, 124540, 3614, 10643, 22800, 2058, 8661, 55964, 55832, 388662, 8925, 221, 1457, 5095, 55567, 6152, 3857, 1376, 3418, 10644, 2898, 498, 8662, 728378, 6812, 137886, 4830, 116085, 55699, 8926, 515, 23066, 5096, 2979, 10381, 84239, 9588, 8663, 1994, 3320, 4831, 51606, 728113, 6813, 157, 2173, 4833, 3921, 6154, 10382, 5493, 6418, 373863, 8664, 4832, 2271, 1832, 6814, 494115, 5478, 6535, 1743, 10631, 2182, 8120, 9045, 155051, 5213, 6138, 645142, 5479, 6536, 30846, 10632, 801, 653598, 362, 118980, 7064, 622, 5214, 2774, 22919, 23450, 5480, 1159, 2199, 119, 3294, 3895, 150160, 4951, 60314, 38, 10236, 1419, 2936, 3017, 4057, 6538, 5481, 5613, 3098, 4820, 10105, 818, 57142, 6141, 7464, 1615, 26227, 6539, 396, 51332, 8653, 10767, 22924, 51199, 2054, 8785, 100506084, 5878, 11031, 6010, 55031, 477, 6142, 29927, 51727, 81855, 285855, 2821, 3861, 217, 6540, 54502, 139728, 1973, 10768, 8654, 6804, 5879, 1632, 51066, 1892, 8125, 1291, 143425, 54505, 7334, 8787, 51069, 127833, 5880, 55033, 4691, 92291, 130872, 7203, 376497, 950, 72, 11034, 5881, 9184, 50804, 23473, 5106, 6295, 8277, 10391, 2941, 6427, 1479, 10657, 116362, 8939, 358, 4241, 57826, 81876, 81, 10128, 83858, 6164, 5239, 6428, 80155, 6560, 23208, 10656, 8676, 9601, 219293, 654231, 1496, 1756, 2664, 4976, 5901, 162605, 1415, 3883, 34, 3184, 25984, 3866, 81611, 654364, 3265, 10130, 27306, 635, 814, 1513, 84517, 375, 283106, 8546, 392,

Supplementary Table 2 (continuation). Ataxin-3 human interactors present in different EvoPPI3 predict interactomes

57699, 1432, 5903, 441531, 10131, 3069, 230, 22949, 84258, 10529, 9736,
553158, 51752, 2890, 9868, 63908, 3329, 79631, 490, 5111, 311, 6300,
22948, 5375, 27044, 6432, 10528, 3849, 140807, 23743, 1807, 328, 7094,
57172, 409, 669, 4190, 10397, 7358, 3312, 6301, 113457, 146754, 10531,
197358, 255626, 390792, 9870, 5906, 5245, 29953, 644591, 2775, 10398,
8416, 159163, 102724023, 9987, 2783, 3303, 5230, 6155, 84245, 55176,
3043, 121506, 27032, 10383, 337, 55968, 1995, 83849, 8665, 4967, 7874,
10120, 79751, 27295, 6156, 1654, 8402, 137362, 55703, 23464, 54514,
27034, 3887, 6950, 10121, 5232, 4147, 793, 1411, 6421, 4703, 30, 3188,
8801, 290, 4705, 9990, 27165, 58480, 11047, 6158, 126393, 4440, 150572,
1509, 8536, 8668, 57555, 126129, 4704, 1769, 57687, 25979, 8802, 140801,
206338, 51084, 908, 9066, 47, 29943, 989, 226, 6687, 27301, 5630,
100287399, 3691, 8803, 6160, 29942, 307, 56896, 94160, 1965, 2894,
144501, 3853, 128, 5499, 9861, 844, 25981, 23603, 7086, 64167, 10257,
5236, 8539, 139741, 2617, 10521, 162, 3316, 51350, 8671, 1104, 5500,
245972, 10787, 1462, 4841, 5105, 2779, 5501, 115827, 7171, 192683, 382,
26330, 54474, 51303, 10606, 6510, 57381, 144983, 116443, 5717, 25934,
9152, 221092, 7038, 445582, 1520, 2040, 57644, 480, 5190, 10211, 55530,
6511, 301, 3843, 221357, 203, 25801, 116444, 5718, 284382, 90019, 821, 58,
3046, 3306, 10212, 7041, 55269, 6512, 1537, 2317, 5719, 57118, 105371242,
220, 2948, 3208, 3728, 10213, 132430, 84467, 79183, 1375, 646817, 2769,
140886, 10873, 9948, 83939, 51700, 3890, 440689, 90550, 27127, 335, 1993,
3191, 84076, 51703, 156, 2752, 1213, 10611, 55930, 2091, 28976, 8761,
51702, 3694, 170487, 4931, 7045, 23165, 254122, 6252, 158067, 196385,
271, 1409, 140625, 408050, 8630, 4666, 84735, 51440, 79716, 2108, 3856,
347273, 6121, 6253, 2897, 201140, 3417, 9688, 2027, 131377, 221223, 8895,
1767, 5313, 10598, 1158, 118, 90141, 10730, 5577, 57505, 1938, 8880, 1499,
5709, 27115, 165324, 340529, 9276, 37, 25794, 28965, 196374, 1418, 3945,
84196, 6237, 2628, 91860, 817, 7295, 378, 6238, 5315, 7429, 10864, 340267,
81554, 286887, 7296, 476, 3310, 55920, 216, 7430, 283848, 6505, 493, 8883,
2773, 3033, 10072, 284904, 102724334, 3911, 314, 166378, 1290, 7431,
6506, 9545, 6638, 7827, 54205, 5713, 25930, 51166, 10998, 9016, 30819,
2935, 7166, 5052, 3276, 81027, 27250, 6507, 55526, 26986, 5714, 71, 10999,
5053, 30818, 11131, 949, 7167, 8224, 28968, 7169, 4218, 3178, 6508, 6640,
23157, 1047, 11000, 11132, 10075, 7168, 23420, 60681, 7170, 57642, 3860,
88, 57510, 6509, 2901, 51167, 3421, 2023, 26063, 148281, 11001, 30820,
29895, 4677, 8773, 728458, 1495, 11151, 6130, 6262, 84617, 127294, 3315,
1414, 1674, 8508, 23176, 33, 5603, 347688, 16, 6263, 23439, 2778, 1073,
197322, 6529, 8775, 1431, 5868, 50, 391, 6000, 342538, 4811, 10096, 6132,
121214, 57531, 79861, 90957, 229, 2940, 10492, 5339, 51056, 6530, 4548,
489, 10890, 212, 353238, 1968, 5869, 9172, 10097, 6133, 125972, 310, 131,
22913, 10493, 6397, 56341, 23708, 6531, 10627, 587, 2923, 3703, 23576,
79731, 11155, 23312, 6134, 408, 847, 3882, 170506, 7455, 2744, 6532,
51719, 83697, 6135, 2906, 23047, 506, 30845, 6533, 523, 653145, 26353,
5872, 57403, 91754, 6136, 1303, 783, 2181, 5211, 10498, 30844, 6534,
83959, 361, 134829, 1759, 85413, 6137, 51060, 3848, 81569, 11011, 27131,
7047, 1491, 6122, 2978, 26999, 23564, 272, 1410, 8632, 9557, 6650, 55671,
6782, 4668, 5859, 2109, 6123, 344905, 90423, 22902, 2028, 6519, 51706,
5594, 1768, 191, 81567, 10881, 2782, 6124, 55013, 4001, 6520, 157807,
4670, 5595, 10880, 2944, 5861, 8239, 1785, 6125, 7314, 5464, 23433, 826,

Supplementary Table 2 (continuation). Ataxin-3 human interactors present in different EvoPPI3 predict interactomes

306, 6389, 5596, 6521, 127, 22907, 4673, 3886, 7051, 4146, 10486, 7447,
6390, 6522, 161, 3187, 4540, 5597, 4674, 7052, 5995, 3609, 7184, 7316,
54749, 23435, 10487, 5598, 10753, 5864, 4675, 29119, 7846, 178, 6128,
7317, 1738, 3690, 30968, 112755, 8904, 1917, 4676, 2893, 9961, 5865,
23569, 10093, 1056, 7054, 1657, 8243, 6129, 23305, 7318, 6261, 3852, 5600,
8639, 26297, 7402, 6345, 283820, 4363, 7534, 2316, 5684, 398, 5816, 26958,
4627, 5948, 202559, 94081, 9119, 388468, 3339, 3859, 219, 10574, 9517,
1374, 3420, 496, 2900, 84172, 4628, 5685, 284217, 8988, 56950, 4892, 5158,
3241, 10443, 10575, 55236, 4200, 4629, 594, 319101, 5686, 29999, 9914,
8857, 10971, 7007, 611, 2171, 10312, 10576, 5687, 10840, 10972, 189, 9124,
5160, 3305, 10313, 7274, 6217, 9520, 3045, 270, 1408, 5688, 378951, 8991,
2947, 7143, 2107, 5161, 7407, 2286, 3646, 368, 4632, 2026, 2768, 5689,
1327, 3028, 125, 84701, 10051, 3889, 11240, 57486, 338785, 9126, 5162,
9390, 102724488, 2670, 3190, 9522, 378949, 5690, 4633, 223, 726, 5956,
284612, 483, 145482, 7277, 53917, 824, 6616, 1962, 378950, 112714, 5691,
4634, 374454, 3181, 4355, 84823, 401409, 26289, 4619, 23382, 5940, 23779,
2052, 10432, 3880, 5413, 55753, 6470, 89845, 2823, 492, 4620, 378948,
54431, 283284, 57470, 2904, 9376, 6205, 201625, 5414, 221184, 1809, 4285,
4621, 345651, 590, 3066, 10170, 2806, 6206, 6472, 2167, 4622, 7529, 9775,
3326, 3927, 22987, 70, 5150, 9114, 509, 948, 6207, 3309, 9512, 58533, 526,
7001, 553115, 2870, 5019, 55757, 6210, 51661, 87, 10570, 80201, 7531,
4624, 364, 1404, 26953, 3032, 3292, 5813, 22989, 606495, 5152, 54039, 445,
3910, 7134, 23520, 653269, 7532, 5550, 80336, 4072, 5682, 4625, 5814,
65009, 23521, 1421, 2934, 51663, 3275, 479, 23519, 7533, 55230, 91433,
4626, 1080, 8854, 11100, 23784, 381, 11232, 9118, 4644, 5701, 94097,
192666, 8343, 6229, 3689, 471, 5568, 8607, 728689, 4645, 55117, 10723,
4513, 5702, 2892, 3071, 488, 211, 5834, 10987, 7155, 83660, 8344, 309,
6230, 3851, 7419, 130, 59345, 10590, 80221, 8608, 6628, 10856, 586, 23541,
548596, 653505, 92906, 5042, 26052, 8345, 407, 667, 2875, 645974,
107987423, 64236, 5704, 151835, 5836, 10989, 10196, 55647, 3915, 8346,
10592, 5573, 10594, 100996331, 522, 5705, 5837, 11254, 23279, 782,
100526737, 343, 8347, 9404, 131890, 2180, 2939, 6233, 4060, 1660, 5706,
360, 1400, 10066, 3101, 8348, 8480, 5575, 23545, 388743, 5707, 1937,
23413, 10067, 6103, 51292, 57502, 539, 36, 3182, 3881, 8349, 9540, 10465,
5576, 5708, 2824, 5972, 23148, 160428, 89872, 2645, 377, 816, 8218, 2905,
1515, 5957, 9260, 85236, 7278, 8467, 10449, 51806, 2125, 50617, 3872, 224,
3693, 4635, 3855, 57223, 6222, 10450, 23530, 825, 64754, 2717, 10846, 126,
4636, 5959, 10055, 7148, 1281, 7280, 5034, 6223, 10319, 3936, 4637, 160,
1102, 51412, 84706, 4771, 10056, 9131, 22871, 51807, 5431, 501, 140465,
2781, 139804, 10982, 56171, 5829, 10057, 7018, 5961, 5036, 8339, 51678,
4000, 79029, 7414, 8735, 10717, 1737, 1315, 5962, 4905, 1656, 8340, 7415,
23534, 8604, 8736, 23270, 225689, 6888, 653781, 10059, 1153, 7284, 8209,
7416, 1413, 8341, 2926, 2747, 5566, 3885, 8473, 10587, 535, 128312, 5700,
5832, 160287, 10985, 10060, 3608, 3868, 8078, 6228, 10324, 7153, 9399,
8342, 5567, 7417, 6624, 1072, 1292, 64714, 3892, 5519, 10540, 57447,
10672, 1731, 6708, 593, 4858, 1650, 2170, 3014, 55728, 9352, 10541, 57446,
6709, 83875, 55860, 1488, 3696, 55335, 1407, 3615, 3875, 4728, 8824,
54938, 6710, 529, 10938, 63923, 51635, 84929, 1066, 3858, 646486, 23358,
2639, 2899, 3419, 6711, 353288, 23491, 10939, 4729, 124, 6184, 83481,
7373, 43, 56922, 3939, 4199, 6712, 10808, 29968, 730262, 140032, 222,

Supplementary Table 2 (continuation). Ataxin-3 human interactors present in different EvoPPI3 predict interactomes

	10015, 26, 1522, 482, 9356, 6185, 205, 823, 10677, 347733, 10016, 60, 3304, 2784, 2319, 643847, 102723728, 158, 54413, 2946, 6187, 8301, 100653049, 2767, 6583, 8565, 102724652, 57835, 1808, 50700, 23217, 83733, 84790, 415117, 112802, 1727, 10135, 3376, 728945, 10399, 9476, 121391, 3018, 4719, 5908, 81617, 10929, 6833, 3099, 22820, 84263, 340990, 85320, 9343, 140679, 525, 4720, 58498, 10137, 785, 8287, 57176, 29958, 86, 5250, 363, 9874, 57175, 2903, 120, 4284, 1160, 5515, 2805, 3065, 39, 4722, 5911, 388951, 6175, 11328, 1420, 51760, 1778, 478, 998, 64577, 27316, 8687, 5516, 10801, 3845, 5912, 4723, 730005, 22824, 3048, 3308, 11331, 24148, 60484, 397, 8688, 57180, 2869, 85453, 55066, 201595, 1974, 218, 5518, 342574, 4329, 2950, 2771, 3291, 495, 1015, 3909, 8689, 10935, 4857, 11199, 51762, 845, 2665, 6857, 57727, 2746, 3185, 4478, 163, 3884, 7385, 9631, 51121, 6594, 5537, 1463, 7915, 10954, 53630, 83759, 125115, 7386, 85477, 1821, 8841, 266629, 55748, 5141, 23244, 3688, 3948, 2810, 10558, 151011, 359, 6860, 91156, 3070, 2631, 2891, 9899, 5935, 2098, 5142, 27339, 51386, 3850, 343263, 203068, 9636, 346562, 10693, 1757, 9900, 5936, 8975, 4747, 6861, 5143, 8050, 7125, 35, 92483, 4012, 8314, 3313, 267020, 10694, 728395, 93408, 145864, 18, 57731, 3735, 7919, 51520, 90634, 5144, 10165, 815, 1075, 7258, 6201, 1514, 2776, 22981, 2597, 55752, 7259, 6202, 2938, 5145, 4354, 8714, 491, 147183, 57468, 6203, 312, 5146, 3888, 6188, 4148, 55869, 6584, 55737, 64061, 3189, 51377, 1459, 79784, 9360, 6189, 10549, 84940, 10681, 4735, 57586, 646480, 2652, 57718, 6717, 9097, 29974, 643834, 64856, 9361, 338, 1996, 4736, 4604, 55342, 1915, 3854, 4738, 55210, 64591, 57060, 6059, 317754, 1314, 143471, 1655, 6191, 23499, 6455, 2895, 9626, 5530, 8701, 55213, 9892, 5928, 10949, 6853, 50853, 3837, 6192, 2192, 1152, 121278, 794, 6456, 1412, 4606, 8570, 5531, 147968, 27068, 811, 3918, 8968, 6854, 10157, 346288, 6193, 144402, 372, 64064, 9628, 91683, 166863, 1510, 1770, 4607, 7514, 2780, 6457, 728403, 3658, 5532, 6855, 119559, 79659, 48, 6194, 5533, 83752, 8572, 9497, 4608, 6856, 728137, 5931, 487, 8970, 7384, 5138, 308, 1267, 9498, 8573, 9630
Predict interactome using <i>D. rerio</i> predict polyQ_22 database using Ensembl orthologies	1509, 2744, 3329, 22949, 5706, 9868, 23279, 1314, 2923, 10694, 23312, 10594, 3735, 23576, 1655, 10528, 22948, 6201, 27044, 5111, 6135, 501, 1152, 1737, 2192, 5707, 644591, 4190, 1412, 140679, 63908, 306, 4060, 160, 3182, 3312, 7358, 10397, 30, 6301, 5211, 6202, 7094, 794, 2875, 6136, 290, 51326, 535, 81617, 5708, 1315, 5576, 5245, 8416, 2664, 6203, 2778, 6137, 5146, 1965, 23743, 3945, 22919, 5709, 47, 811, 372, 1413, 10598, 307, 5478, 10531, 3376, 5213, 10399, 2746, 1738, 6138, 9276, 9045, 226, 10135, 93408, 131890, 6164, 7155, 23176, 8939, 5834, 81611, 51386, 5702, 191, 2271, 7386, 6428, 7419, 2027, 6230, 6132, 27339, 160428, 4976, 102724652, 3688, 5901, 2775, 4811, 653145, 5868, 10856, 57727, 158, 223, 483, 10657, 2109, 4513, 51121, 728689, 10492, 340529, 408050, 728458, 1491, 1995, 10097, 653505, 10130, 23244, 9172, 10989, 51520, 55748, 5704, 338, 10890, 645974, 9601, 6628, 6529, 2597, 10493, 1654, 347733, 6133, 825, 10131, 23047, 83858, 11155, 81876, 5837, 22981, 4747, 5705, 9900, 10693, 100996709, 3313, 3703, 3069, 7125, 6134, 6193, 7316, 10157, 6160, 7184, 203068, 1757, 2098, 5995, 347273, 522, 782, 3837, 18, 4905, 7052, 587, 847, 29943, 8802, 4673, 140801, 9892, 4607, 8736, 1968, 7514, 506, 3609, 8604, 10487, 7415, 51678, 7284, 6194, 6128, 1660, 23569, 50, 10059, 7846, 9861, 5864, 6888, 6855, 30968, 8803, 4608, 489, 5533, 5566, 8671, 8473, 34, 8539, 229, 7416, 51084, 6228, 6129, 8243, 10324, 8078, 523, 9961, 23603, 5832,

Supplementary Table 2 (continuation). Ataxin-3 human interactors present in different EvoPPI3 predict interactomes

	8904, 10985, 80221, 5567, 6624, 112755, 8639, 328, 1400, 3608, 7417, 2652, 7318, 80155, 2782, 6261, 7384, 5106, 6857, 311, 23439, 283820, 4677, 5701, 5700, 101060233, 10787, 5568, 4478, 9631, 8607, 539, 1303, 6427, 1937, 7385, 6295, 6229, 8402, 16, 471, 6189, 6222, 6156, 845, 9987, 6717, 2665, 26999, 1072, 84245, 1462, 2779, 5594, 1267, 9361, 55737, 48, 6223, 55671, 5232, 7280, 9097, 10121, 6124, 1414, 1479, 4001, 5034, 178, 308, 5959, 6950, 55968, 10055, 7874, 4703, 51377, 4736, 4670, 5595, 1495, 1674, 2747, 162, 6520, 6421, 130162, 6191, 6125, 346288, 2617, 6158, 8239, 143471, 6059, 11047, 64064, 407, 9131, 56896, 10056, 54749, 4704, 2780, 9626, 126129, 635, 8668, 5431, 5464, 6389, 23435, 5036, 5961, 7051, 309, 5928, 23534, 8801, 4705, 90423, 33, 4146, 8570, 4540, 1431, 7414, 163, 6456, 6390, 7447, 3646, 100506084, 396, 7407, 29968, 217, 5161, 6185, 5095, 526, 6152, 84076, 51440, 4830, 90550, 5689, 6812, 8662, 4632, 83481, 9522, 3418, 55832, 818, 730262, 11240, 6252, 5162, 1160, 2639, 5096, 7045, 9126, 150572, 477, 4666, 5690, 1615, 3190, 3320, 38, 8663, 6418, 7277, 1892, 6187, 344905, 6121, 6154, 998, 1632, 81567, 3046, 381, 5691, 3417, 3921, 6616, 8565, 5493, 8664, 23165, 9688, 10449, 3030, 23530, 23464, 10383, 5230, 6155, 6188, 51606, 6122, 120, 5957, 5956, 4148, 79716, 728113, 218, 1291, 22902, 4668, 8665, 3843, 10574, 56886, 1499, 5223, 26958, 1759, 9352, 26330, 3339, 7072, 5685, 201140, 5718, 4070, 4200, 8659, 3420, 10575, 491, 5224, 5091, 3306, 10211, 1938, 57447, 9914, 1808, 2052, 5719, 4728, 1743, 3192, 10938, 5686, 84239, 10576, 51702, 10213, 2670, 3712, 6017, 10212, 314, 55699, 509, 4729, 2784, 10972, 10939, 3728, 6711, 10873, 3939, 10643, 3614, 3028, 3419, 7373, 5358, 9520, 10313, 6217, 5160, 6184, 5094, 2947, 51602, 8991, 37, 476, 1809, 2752, 3191, 8661, 144983, 2805, 4723, 29927, 270, 3065, 335, 5714, 4624, 6507, 10570, 26227, 7531, 3276, 4284, 8125, 267020, 5053, 441531, 189, 7167, 10999, 3098, 220988, 1213, 1278, 2058, 29895, 3033, 51166, 302, 8787, 4072, 5682, 6640, 5550, 7532, 3178, 3308, 2107, 23420, 611, 58533, 221, 7168, 10075, 23519, 481, 5814, 1376, 2091, 7533, 2806, 24148, 4329, 515, 1327, 1522, 9118, 11331, 23157, 10935, 55860, 4857, 255626, 2026, 4627, 368, 498, 1409, 2319, 5684, 43, 5519, 10606, 3421, 6345, 10540, 81855, 482, 2108, 2628, 7203, 2823, 23521, 5717, 1994, 5908, 155051, 2023, 653598, 203, 4719, 10929, 89, 5479, 593, 645142, 10632, 6205, 9343, 84823, 1015, 55920, 7064, 2316, 1666, 2771, 4720, 219, 728945, 1292, 3895, 6206, 8287, 10236, 353238, 1374, 3326, 9775, 204, 7529, 6538, 7430, 6472, 529, 1650, 4057, 23450, 6207, 6141, 55526, 9114, 10072, 3032, 60314, 5713, 4722, 301, 821, 8653, 6506, 84263, 1537, 7431, 2821, 11328, 3309, 5250, 6175, 6142, 5019, 5052, 6010, 480
Predict interactome using <i>D. melanogaster</i> modifiers genes using DIOPT orthologies	1277, 8886, 57510, 79048, 5583, 8622, 5451, 838, 1196, 55660, 60681, 7169, 3276, 654346, 5055, 30820, 6905, 408187, 659, 5584, 5452, 402635, 55529, 7170, 5981, 79576, 6906, 2901, 821, 1601, 41, 8884, 57508, 5581, 5449, 90673, 1520, 1780, 7431, 144455, 29761, 11131, 30818, 57111, 382, 9149, 574537, 1763, 51035, 5582, 8621, 9282, 24, 3958, 81027, 7168, 56980, 5054, 30819, 10207, 11132, 5979, 9939, 1409, 283987, 102724652, 112869, 23284, 9675, 10600, 5579, 51692, 101930420, 55656, 94239, 5315, 5976, 1767, 10204, 140890, 8883, 26985, 392862, 4171, 6769, 5580, 6637, 246329, 107984138, 196385, 83667, 10733, 92, 648791, 6241, 54600, 30817, 3293, 10730, 57636, 5709, 89876, 168507, 2952, 9805, 2334, 5313, 54729, 115509, 8484, 339745, 9145, 51029, 27115, 11127, 6899, 83401, 5710, 51427, 57505, 5578, 1375, 81551, 3310, 94104, 146310, 9542, 440689, 51030, 8089, 25926,

Supplementary Table 2 (continuation). Ataxin-3 human interactors present in different EvoPPI3 predict interactomes

79568, 10992, 254394, 3621, 8085, 11124, 27244, 57634, 2321, 6500, 55916, 1460, 25791, 29094, 5972, 7029, 157807, 9143, 6897, 5708, 23413, 124359, 842, 9804, 440435, 23546, 3817, 29888, 9144, 10201, 25792, 10333, 123169, 5705, 89872, 23410, 10594, 338692, 3084, 5441, 728597, 2905, 825, 8348, 23279, 7027, 6895, 10991, 57633, 5706, 158866, 23411, 145389, 3182, 100506164, 6499, 3003, 727, 8349, 9138, 162699, 5174, 23144, 3233, 9931, 2892, 29883, 23673, 8346, 1072, 9535, 27241, 83659, 10196, 4175, 94097, 653689, 5704, 79695, 284129, 10725, 8347, 25920, 6233, 55120, 341208, 5172, 9136, 10193, 8343, 83658, 204851, 3297, 7022, 4776, 5701, 9797, 79034, 59345, 8344, 84451, 10194, 55515, 83657, 55911, 79693, 143279, 10723, 51022, 59082, 5702, 339488, 2875, 221895, 5438, 8345, 64499, 80222, 23404, 8341, 1025, 7020, 2764, 9134, 10720, 162968, 90527, 165082, 26047, 9927, 4642, 5699, 23008, 51283, 79955, 8342, 54456, 7021, 283208, 10060, 5700, 4775, 4643, 285190, 58155, 1106, 5436, 54455, 27237, 8339, 79954, 643394, 7150, 10057, 30009, 4772, 349667, 4640, 5433, 9397, 400668, 8340, 5037, 9133, 4773, 8737, 81539, 4641, 29085, 5698, 56171, 2487, 1609, 292, 9662, 128312, 7280, 51147, 10055, 387990, 377, 5695, 3156, 80349, 1515, 10716, 440836, 246213, 718, 54980, 134829, 85236, 51809, 201158, 19, 6885, 56301, 23269, 84708, 168391, 4000, 79157, 360, 2261, 3301, 51013, 1660, 160728, 100534599, 6882, 23002, 2, 701, 50484, 10846, 23266, 92105, 8468, 9129, 25777, 347688, 5033, 342667, 23135, 10054, 343, 4637, 5694, 80346, 135228, 170506, 10847, 84706, 126326, 283871, 55900, 56957, 5558, 79816, 5294, 3727, 1387, 6880, 9126, 94086, 2589, 4634, 5691, 54578, 150160, 222698, 22868, 79947, 51012, 222962, 29078, 57486, 26040, 27097, 6881, 143686, 3305, 51144, 6749, 23397, 4635, 54579, 54973, 6613, 2312, 51802, 752, 55898, 30000, 53, 7274, 3791, 10049, 118924, 256394, 833, 339105, 134, 9785, 4632, 5689, 54576, 51537, 283742, 5557, 5293, 390502, 22998, 729540, 7407, 55238, 10446, 153201, 10050, 7936, 5822, 9918, 54577, 57616, 3190, 123283, 23788, 1080, 1600, 5686, 2900, 22863, 40, 7404, 5422, 133060, 5290, 10179, 8065, 5158, 9915, 381, 3778, 6612, 4038, 10576, 5687, 23657, 56954, 55236, 84435, 462, 8462, 5291, 3177, 81793, 5159, 54575, 6877, 143162, 94081, 5688, 5684, 9648, 340562, 6609, 146862, 364, 5288, 7534, 4363, 5156, 220988, 10309, 203547, 6874, 23522, 8988, 3842, 57084, 5685, 390110, 23787, 54970, 29072, 51270, 139596, 55366, 5289, 10574, 9517, 7403, 10178, 3241, 100509620, 84698, 10971, 8989, 6875, 51069, 10769, 6541, 51201, 124930, 27285, 123, 1521, 1781, 9184, 11034, 221188, 1244, 25833, 10902, 6542, 23190, 26097, 79080, 4296, 26229, 79873, 5089, 54505, 80005, 8260, 7071, 100533496, 81855, 129685, 27151, 366, 4293, 7332, 3925, 65997, 7993, 6936, 129684, 29926, 54898, 92292, 3064, 9843, 197407, 54107, 8654, 55296, 55692, 4294, 7201, 23321, 92291, 7994, 5087, 55031, 29927, 54899, 51728, 55293, 25960, 647174, 6801, 118980, 2953, 54764, 55689, 51593, 515, 11162, 165829, 3912, 28999, 10237, 3652, 1116, 8916, 9973, 6670, 123207, 89778, 2693, 55690, 152485, 29925, 1278, 2058, 4817, 85417, 125981, 89777, 3017, 146540, 6667, 728577, 1457, 3797, 284613, 84889, 5346, 102724159, 5214, 29922, 7064, 5082, 140, 26091, 8914, 4172, 340371, 839, 5611, 23581, 1977, 7329, 5347, 88455, 9311, 55289, 79733, 23049, 11157, 81847, 9704, 5608, 146542, 65991, 219743, 7458, 253832, 2906, 7062, 80789, 140465, 988, 55290, 26353, 84885, 7855, 23314, 55554, 83696, 10498, 79072, 146279, 5213, 254359, 25825, 7327, 28996, 25957, 79731, 890, 1410, 7984, 2710,

Supplementary Table 2 (continuation). Ataxin-3 human interactors present in different EvoPPI3 predict interactomes

101927685, 8777, 8909, 6795, 317649, 9570, 5606, 8513, 25822, 7324, 389396, 1768, 26086, 79071, 339318, 85015, 93, 25823, 51191, 140732, 7325, 8382, 5211, 6132, 7189, 23309, 83693, 3298, 9039, 55681, 64005, 5604, 9568, 3217, 3916, 56341, 140733, 29123, 7322, 26084, 51454, 146547, 84617, 2535, 6926, 63874, 65988, 1120, 8908, 5605, 3315, 55285, 79993, 28992, 29124, 26085, 7323, 5205, 10226, 90826, 3622, 7187, 54490, 2322, 4677, 10755, 25818, 285282, 84220, 57396, 1201, 7188, 90957, 54623, 11284, 51320, 3899, 6924, 4176, 100529257, 6528, 1981, 1103, 144568, 7321, 7317, 54620, 11281, 7185, 64926, 10753, 27005, 7318, 23438, 7186, 7979, 122402, 6790, 9829, 293, 5469, 79064, 10622, 117248, 7447, 2893, 29777, 7976, 253314, 114, 404203, 81573, 22907, 55146, 1073, 2015, 9694, 51314, 7316, 7184, 54619, 388743, 16, 276, 79722, 79590, 8506, 4542, 1674, 5463, 4001, 10484, 361, 9295, 23037, 3920, 6785, 2782, 3221, 8767, 7314, 84873, 65055, 340654, 79852, 5861, 342372, 8900, 283576, 9425, 50650, 25809, 7311, 10350, 79587, 147746, 11143, 60559, 1384, 5858, 55274, 10483, 5462, 164395, 6519, 81832, 256356, 10351, 25942, 56993, 3204, 79982, 161753, 100287226, 22904, 11012, 229, 6652, 6520, 8766, 5591, 10612, 1793, 5459, 283446, 79849, 11273, 149069, 642843, 7177, 55008, 57122, 27129, 51308, 834, 6913, 81831, 1955, 5592, 83549, 51440, 5460, 135, 11274, 10349, 25940, 55669, 5725, 6782, 23430, 8632, 2492, 5593, 6650, 56063, 64123, 2897, 1337, 5457, 3417, 9289, 25937, 79979, 898, 57120, 116447, 54477, 378, 6911, 2019, 84206, 4035, 5590, 2475, 84074, 5458, 10347, 51043, 27127, 20, 22899, 165721, 10083, 280, 54346, 6780, 23428, 2965, 29899, 5719, 1665, 284904, 55268, 5455, 149998, 23293, 57646, 5984, 10212, 51304, 6909, 23557, 158983, 5588, 1047, 5456, 146167, 3843, 29107, 116444, 79183, 10477, 27125, 64518, 166647, 6117, 57647, 4928, 5985, 79844, 6910, 83940, 1486, 23426, 5585, 1729, 689, 5453, 116443, 27254, 3988, 57512, 136259, 7171, 5982, 83939, 2590, 149076, 5586, 5454, 11136, 256364, 23160, 90809, 3306, 23556, 8890, 5983, 780776, 114789, 5253, 150465, 7367, 84135, 51103, 269, 2805, 3065, 5650, 22827, 360023, 7368, 55331, 968, 51631, 57709, 6311, 81624, 132625, 11068, 206358, 22824, 347734, 81622, 84265, 60481, 7365, 23353, 4284, 7101, 29959, 347733, 10933, 64446, 3308, 7366, 63917, 54536, 51497, 127602, 51761, 57176, 282996, 85320, 27314, 3018, 1198, 7363, 29957, 219970, 9213, 25861, 347741, 7099, 347475, 51099, 27315, 5646, 84658, 222611, 840, 5514, 7364, 10403, 116369, 284451, 140032, 327657, 7100, 10139, 3815, 2937, 8021, 51096, 6568, 7757, 55852, 6304, 23481, 7097, 2903, 823, 6833, 5644, 55589, 58496, 3960, 102723796, 9344, 9476, 399664, 6041, 90338, 9212, 84128, 23214, 79634, 24139, 222484, 55718, 111, 2890, 63910, 5509, 286436, 2272, 1769, 10002, 147699, 533, 4173, 23479, 143471, 114780, 2289, 8417, 100101267, 354, 51755, 8549, 144132, 7096, 10131, 129049, 3035, 51224, 285641, 5903, 83860, 1215, 10924, 54659, 10528, 4582, 148229, 3214, 8678, 10660, 5507, 147965, 10396, 405753, 3913, 203062, 643847, 79895, 8811, 128387, 3832, 64834, 84916, 100130733, 728929, 440503, 10529, 85445, 3312, 347088, 81876, 84651, 221035, 7090, 23210, 139378, 5901, 54657, 55714, 57167, 284194, 100289279, 9337, 64434, 10526, 405, 7091, 3201, 222487, 5902, 54790, 51619, 54658, 147968, 5506, 1982, 340481, 226, 9470, 10527, 6299, 7088, 27436, 135152, 51616, 167127, 729857, 200424, 653489, 6560, 9863, 23076, 203068, 10524, 55315, 83855, 92840, 89801, 7089, 1509, 3964, 54788, 23473, 84912, 92312, 9732,

Supplementary Table 2 (continuation). Ataxin-3 human interactors present in different EvoPPI3 predict interactomes

2745, 23077, 1607, 3184, 6297, 85441, 10525, 283537, 10257, 84779, 2894, 94160, 10786, 8672, 6690, 25981, 59277, 375, 6558, 1513, 84515, 165545, 2992, 277, 2033, 51747, 9730, 340485, 196463, 358, 6691, 976, 6559, 10655, 6423, 150483, 9462, 8405, 2958, 3299, 93233, 9991, 51479, 699, 9595, 6424, 51347, 10256, 84908, 93100, 57030, 79887, 114769, 9728, 64429, 79623, 27037, 8539, 83719, 89797, 3316, 6557, 6425, 256979, 150353, 5232, 80150, 129831, 4836, 6950, 101928917, 10650, 89796, 3725, 2587, 4440, 8668, 6422, 51611, 153129, 29941, 8140, 3303, 4705, 9990, 11047, 230, 6687, 89795, 5494, 23464, 126526, 51, 5230, 10383, 9987, 27164, 126129, 23200, 8798, 26503, 84641, 126393, 5495, 84243, 57026, 4245, 387921, 55176, 285126, 8667, 4703, 5496, 2898, 3418, 56891, 55173, 22933, 10249, 50945, 10381, 57551, 93624, 134187, 3337, 2020, 54512, 51473, 7475, 201625, 4036, 720, 10382, 21, 64423, 58477, 23331, 124540, 10643, 2962, 83447, 132864, 362, 27159, 79616, 55567, 2263, 2783, 23064, 23460, 3222, 2280, 6416, 3320, 8662, 10644, 7341, 10248, 23197, 51340, 55304, 27293, 166, 7738, 23461, 63892, 29800, 126520, 6545, 1730, 57019, 5356, 51205, 4299, 79085, 80142, 27290, 64682, 9056, 6017, 100130086, 161725, 131540, 3030, 10642, 53981, 149041, 22798, 9585, 401409, 5225, 7471, 55170, 10114, 9057, 10246, 84634, 260425, 10907, 79613, 51602, 100529239, 83443, 1794, 55035, 86614, 55299, 7468, 835, 5222, 55, 5090, 124535, 84897, 6676, 50940, 136, 54904, 4430, 1436, 7337, 8526, 7469, 10112, 131405, 10244, 57546, 23589, 29799, 1992, 6871, 27087, 5682, 22859, 9646, 56288, 7532, 8193, 5286, 8325, 63951, 64083, 25766, 54834, 1374, 8722, 6608, 22992, 9647, 7533, 80335, 3309, 5287, 113235, 8326, 56949, 54967, 79675, 10969, 55760, 4008, 25896, 26953, 285203, 9908, 57211, 23253, 10569, 3015, 8323, 837, 1195, 54700, 51661, 55625, 5152, 7266, 3275, 81788, 1178, 90507, 7531, 80333, 1975, 79937, 8324, 63950, 10963, 154664, 89845, 120103, 7528, 57473, 100526693, 8321, 63947, 54830, 3603, 4285, 8453, 89978, 6339, 387082, 5150, 388403, 9246, 824, 1263, 2904, 726, 1962, 441381, 84823, 9643, 7529, 6472, 23120, 219539, 6340, 8322, 23517, 55756, 8454, 55624, 63948, 5151, 54963, 79935, 285343, 368, 201176, 139735, 266977, 23381, 5544, 3927, 6601, 8451, 2009, 64077, 6337, 4221, 11226, 5016, 2286, 2806, 3066, 3326, 9905, 219541, 100652871, 5545, 23250, 440854, 6470, 345062, 8320, 63946, 157310, 727851, 84954, 6338, 8452, 11227, 64210, 26154, 79799, 139341, 8449, 6599, 3215, 7128, 54033, 3914, 5146, 84687, 1118, 339122, 144363, 8714, 221393, 8450, 3053, 6204, 57205, 65265, 1297, 142910, 3313, 79665, 5143, 2580, 58524, 55617, 6597, 128553, 5408, 203427, 1459, 56938, 5276, 5144, 22849, 57468, 137362, 4174, 221527, 6862, 144233, 7919, 841, 221395, 643418, 54957, 6598, 50861, 10694, 54825, 57864, 3816, 5145, 55089, 84289, 5141, 27339, 164714, 10294, 8841, 284293, 6595, 7652, 828, 5406, 8313, 8445, 5274, 5142, 440068, 9238, 1510, 57466, 3965, 8974, 51520, 84287, 117583, 64072, 8710, 9371, 26019, 8314, 291, 5407, 22848, 5275, 7253, 54556, 5271, 892, 8310, 23241, 2891, 112, 84946, 643414, 57198, 64333, 23770, 2712, 2972, 9368, 5272, 1770, 2030, 2550, 79660, 27074, 8312, 5273, 10558, 85474, 153090, 2810, 8444, 5269, 160622, 7251, 84680, 359, 2260, 3300, 79659, 9101, 1659, 51118, 114805, 8837, 22843, 11216, 5270, 10423, 8309, 5138, 23636, 8970, 100996928, 159963, 57461, 254173, 6192, 5267, 2324, 9099, 79657, 80714, 9231, 25879, 400966, 114803, 113746, 283106, 255626, 57194, 27069, 1642, 7514, 84678, 100528030, 102724642, 9496, 43849, 51248, 5268, 51646, 374407, 4343,

Supplementary Table 2 (continuation). Ataxin-3 human interactors present in different EvoPPI3 predict interactomes

	6193, 406, 3202, 119559, 286410, 84148, 29844, 79788, 146, 1983, 1105, 55345, 441925, 107983955, 43847, 9493, 11211, 10947, 27067, 51114, 92736, 8701, 7644, 23500, 8437, 11080, 6191, 9230, 6059, 26011, 146754, 9891, 79654, 55607, 80315, 55869, 2895, 9095, 26140, 60493, 100506888, 1514, 10681, 60625, 8831, 23233, 1075, 55870, 54681, 278, 5132, 9096, 5000, 347853, 160762, 26009, 157855, 4736, 8568, 51377, 54942, 390010, 54810, 6318, 55206, 363, 643834, 257218, 57452, 157724, 11075, 140679, 2264, 2784, 58509, 80312, 8697, 883, 57584, 10943, 65117, 100506084, 5394, 2281, 653082, 866, 57057, 64324, 84671, 7905, 51639, 5791, 51507, 9887, 4602, 133690, 4598, 5655, 687, 7637, 55204, 85462, 91408, 8034, 134218, 9223, 5788, 128272, 3726, 54941, 57055, 11338, 6317, 29968, 284040, 1629, 85329, 414061, 10413, 80705, 3304, 5921, 10149, 57451, 58508, 151649, 133688, 1795, 5653, 10806, 92595, 85460, 7371, 10146, 836, 345222, 55466, 10939, 55862, 56919, 5786, 9618, 3973, 80306, 29966, 3632, 1356, 55599, 54542, 53353, 64848, 10808, 54674, 22828, 115584, 1859, 2899, 119548, 1339, 57050, 284439, 93650, 54407, 56257, 60485, 83741, 9351, 1778, 10276, 11333, 23621, 26528, 65109, 404281, 721, 283383, 1761, 8427, 55333, 23093, 11202, 51501, 8295, 64978, 65110
Predict interactome using <i>D. melanogaster</i> modifiers genes using Ensembl orthologies	27241, 2875, 5902, 54658, 10725, 9535, 130502, 54659, 84916, 122706, 9337, 29883, 10196, 7091, 26119, 3297, 5705, 55714, 54657, 123169, 51222, 2892, 10660, 8678, 10527, 226, 23542, 9470, 25920, 55120, 7089, 83659, 10194, 8939, 2289, 10723, 5702, 6560, 147746, 340654, 79034, 64434, 1460, 7090, 23210, 1607, 64499, 23144, 5901, 51616, 140679, 80222, 92312, 51022, 59082, 100287226, 5438, 10525, 54788, 84450, 6297, 163933, 256356, 533, 9804, 81551, 347853, 5708, 51029, 54332, 842, 23479, 57636, 643834, 9143, 7029, 4173, 5972, 8880, 5709, 79634, 9805, 79568, 253314, 7757, 10730, 94, 23546, 2890, 25792, 10201, 29888, 10991, 2937, 53405, 8811, 727, 5706, 23279, 64834, 5441, 1198, 10594, 55718, 283106, 7027, 55916, 2905, 3035, 9869, 89872, 102724652, 51224, 23214, 27244, 6500, 3832, 3312, 30811, 8085, 51611, 84908, 8140, 23269, 83983, 79887, 11047, 283237, 81539, 93100, 51479, 279, 718, 9990, 29941, 5695, 8668, 10716, 90655, 10650, 146547, 6422, 574537, 9991, 10057, 4705, 5696, 54980, 4211, 80150, 1238, 6423, 64427, 360, 56301, 230, 9462, 2992, 4245, 57026, 9129, 343, 10054, 6882, 23002, 5693, 2, 3156, 3725, 2033, 8468, 701, 5232, 10055, 25778, 6950, 4836, 55176, 4637, 5694, 8667, 84706, 153090, 5562, 79157, 144568, 23135, 140465, 79693, 54456, 7021, 84912, 10060, 144363, 277, 2487, 10720, 6558, 219743, 126129, 9136, 699, 7088, 54455, 119391, 1025, 3299, 7022, 23076, 94160, 2764, 2894, 126526, 10655, 6559, 9730, 142910, 5436, 358, 58155, 5037, 1072, 30009, 129831, 6424, 253832, 9397, 137362, 83852, 10257, 79954, 7020, 9927, 157777, 100506164, 93233, 5699, 2958, 23404, 5698, 9728, 84446, 6557, 8539, 114769, 6425, 29085, 3316, 27037, 5456, 29966, 6117, 81629, 10146, 29107, 3988, 6909, 836, 27125, 55204, 134218, 140890, 1339, 23293, 2965, 133690, 25937, 285641, 9024, 4928, 29108, 140032, 6910, 10808, 1047, 55599, 4598, 54542, 221687, 5586, 80306, 3973, 23556, 83940, 721, 8295, 79844, 3843, 1192, 83741, 26528, 347475, 57050, 65110, 23557, 7371, 56257, 883, 1859, 93650, 2899, 3306, 10212, 90809, 11136, 5983, 643414, 4216, 51501, 9022, 363, 5719, 8890, 283987, 79982, 5394, 23430, 10349, 57057, 6913, 55274, 5922, 27129, 83549, 25940, 135, 834, 1337, 60559, 2475, 3726, 60625, 8697, 3417, 5460, 54810, 119548, 116443, 55669, 10350, 9095, 51507, 57122, 51639, 54942,

Supplementary Table 2 (continuation). Ataxin-3 human interactors present in different EvoPPI3 predict interactomes

7905, 64324, 5725, 84671, 2492, 10746, 6449, 5458, 116444, 10347, 10413, 23428, 1629, 2019, 280, 20, 11338, 29968, 27127, 6911, 10941, 5590, 58509, 441925, 133688, 2897, 11273, 7177, 91408, 11075, 3304, 54346, 643418, 361, 58508, 206358, 388743, 5459, 57451, 347734, 51099, 6637, 132954, 780776, 5514, 7364, 5316, 333926, 840, 27315, 10204, 57508, 94239, 146167, 5976, 1409, 22824, 5581, 29959, 9479, 85320, 345222, 7365, 6241, 54600, 10139, 6833, 2805, 22821, 5710, 1375, 25926, 5578, 9476, 5313, 23481, 55656, 1196, 6304, 2334, 83667, 9145, 6899, 9675, 5579, 2903, 51030, 159119, 27314, 7363, 153768, 5315, 3310, 94104, 89876, 51427, 90338, 9939, 65109, 5584, 269, 56917, 51103, 5452, 7368, 25932, 22827, 7169, 54407, 838, 60485, 6905, 4857, 157724, 5585, 821, 5453, 23621, 7171, 84135, 11333, 79048, 22828, 55529, 7170, 3632, 5981, 51631, 11068, 27254, 100533496, 284439, 51035, 401236, 5582, 102724428, 1763, 24, 7366, 4284, 60681, 7101, 9149, 112869, 10933, 55660, 149998, 5583, 2901, 7367, 1601, 5253, 9283, 132160, 7168, 3308, 54536, 9282, 165721, 5979, 91862, 151649, 8886, 51497, 57466, 51454, 92745, 51520, 1770, 6926, 26019, 63874, 1510, 153129, 128553, 54758, 5406, 1201, 5142, 139189, 83693, 55285, 168391, 26086, 8909, 201158, 5407, 9371, 3313, 79993, 6924, 222962, 285126, 54623, 51320, 54557, 222698, 23770, 100293534, 2972, 11284, 54490, 3298, 60436, 84946, 135228, 643596, 7652, 23309, 27339, 6595, 100130086, 1103, 2891, 8444, 3899, 10558, 85474, 283871, 7189, 988, 7855, 28996, 390110, 4750, 25957, 23314, 8714, 26353, 283208, 1459, 162968, 64077, 65266, 4817, 2906, 5213, 23381, 196463, 57007, 25825, 65265, 81847, 79799, 83696, 6862, 101928917, 6598, 10694, 57468, 340485, 841, 8513, 84290, 22849, 54033, 80789, 84885, 25822, 100271927, 1410, 23049, 26154, 150353, 2580, 63876, 50861, 57864, 160857, 5211, 25823, 55089, 83893, 54957, 11157, 6191, 7314, 79788, 6059, 79590, 1514, 7050, 5861, 9891, 131583, 55146, 7644, 8437, 9231, 6192, 10156, 7976, 100534599, 29777, 340481, 25879, 80315, 284129, 147968, 167127, 8570, 26011, 10351, 54681, 5132, 56993, 122402, 5000, 55870, 278, 8831, 1659, 25942, 57060, 6519, 10483, 5462, 51114, 9295, 204851, 11211, 10947, 339488, 2895, 22904, 26009, 8766, 9757, 4212, 6652, 51377, 8568, 5463, 6520, 84277, 229, 359, 10484, 11216, 143471, 1642, 57396, 2861, 23438, 1073, 10423, 147699, 27005, 8309, 23241, 57461, 64926, 2829, 10953, 10755, 84680, 4677, 65125, 5469, 3315, 8310, 79064, 7253, 6193, 51646, 16, 145389, 57460, 158866, 276, 147965, 51118, 2145, 9694, 9496, 7514, 2535, 387082, 27069, 78997, 81573, 3300, 80714, 1983, 2893, 79657, 8837, 7251, 23657, 10642, 643847, 5291, 5225, 79813, 5093, 53981, 1193, 9056, 118924, 317649, 381, 51537, 54576, 219541, 6612, 80142, 2900, 3030, 400793, 364, 23460, 55, 1600, 5094, 30000, 57019, 3305, 90316, 132864, 79613, 9057, 4215, 10907, 285343, 94081, 5688, 54575, 6613, 10643, 54970, 23589, 7403, 7469, 10574, 1387, 23787, 9253, 5090, 8988, 64682, 84897, 835, 6874, 134187, 260434, 10970, 51205, 136, 6545, 7404, 5422, 29800, 126326, 4299, 5356, 22863, 10179, 23788, 55035, 57546, 6875, 83443, 51270, 86614, 5686, 157310, 340371, 124535, 92105, 137814, 23397, 720, 50484, 21, 29078, 6880, 10910, 4634, 84243, 51013, 7475, 10383, 5230, 50945, 54579, 94086, 284613, 3303, 2898, 53, 79816, 6881, 26040, 27097, 4635, 51012, 362, 9722, 7407, 79947, 5293, 405753, 1515, 550631, 10248, 653689, 3337, 10049, 23064, 55173, 199720, 833, 5689, 54578, 134, 4632, 54512, 3727, 5557, 10644, 8662, 55898, 63892, 5294, 123283, 2962, 54973, 154664, 9126, 51340, 22933,

Supplementary Table 2 (continuation). Ataxin-3 human interactors present in different EvoPPI3 predict interactomes

	55238, 5822, 9918, 54577, 3190, 57616, 23197, 5558, 83447, 166, 3320, 3066, 3326, 29925, 54039, 9643, 149041, 92291, 6472, 90507, 3797, 3927, 4285, 55625, 89978, 1457, 51661, 1977, 8321, 9246, 515, 9973, 54700, 2904, 25896, 146862, 26953, 368, 129685, 57473, 10569, 9446, 3309, 89845, 285203, 23517, 9908, 55756, 8914, 140, 203547, 8452, 6470, 4221, 5347, 839, 5214, 5082, 51593, 54764, 84889, 6801, 2058, 65267, 54102, 7528, 8320, 7329, 23581, 64210, 9647, 150094, 3064, 8722, 448834, 1244, 6542, 1374, 7533, 5287, 1195, 8260, 8326, 143686, 9184, 3925, 1975, 55299, 11034, 10902, 284076, 51069, 54967, 7468, 56949, 366, 4363, 51201, 4297, 5288, 5222, 125981, 5089, 79873, 81855, 10969, 64083, 7531, 153201, 29927, 80335, 55296, 387990, 8324, 79080, 339105, 283742, 81788, 837, 54834, 6936, 7993, 120103, 1080, 55031, 6541, 7532, 25766, 93349, 5286, 8325, 5087, 3015, 6871, 55692
Predict interactome using <i>C. elegans</i> modifiers genes using DIOPT orthologies	5286, 2309, 5287, 2308, 5288, 3480, 3643, 5289, 5290, 3645, 5291, 5293, 5294, 4303, 100132074
Predict interactome using <i>C. elegans</i> modifiers genes using Ensembl orthologies	4287, 5287, 5286, 5288, 5294, 5291, 5293

Supplementary Table 3. Ataxin-3 human interactors present in main databases, excluding the 151 in Sousa e Silva et al. [106], that are also expressed in reference SCA3 tissues, present in human polyQ database (N=252) in different EvoPPI3 predict interactomes databases

Predict interactomes databases	GeneID
A- <i>D. melanogaster</i> modifiers genes and <i>D. rerio</i> predict polyQ_22 databases	<u>7531</u> , <u>10574</u> , <u>5901</u> , <u>51520</u> , <u>3309</u> , <u>515</u> , <u>5719</u> , <u>226</u> , <u>5689</u> , <u>5686</u> , <u>5702</u> , <u>10694</u> , <u>2058</u> , <u>10383</u> , <u>1072</u> , <u>3326</u> , <u>6191</u> , <u>7514</u> , <u>6950</u> , <u>29927</u> , <u>5706</u> , <u>5709</u> , <u>10594</u> , <u>7407</u> , <u>5708</u> , <u>29968</u> , <u>3313</u> , <u>5230</u> , <u>3190</u> , <u>3310</u> , <u>3305</u> , <u>3315</u> , <u>5315</u> , <u>3304</u> , <u>5701</u> , <u>5700</u> , <u>3276</u> , <u>5682</u> , <u>5685</u> , <u>5704</u> , <u>10576</u> , <u>7431</u> , <u>7184</u> , <u>10382</u> , <u>1778</u> , <u>5093</u> , <u>5687</u> , <u>347688</u>
<i>D. melanogaster</i> modifiers genes using DIOPT and Ensembl orthologies not present in A	<u>10527</u> , <u>23197</u> , <u>8678</u> , <u>5902</u> , 1629, 79568, <u>55660</u> , 5694, <u>51639</u> , <u>6117</u> , 840, 3064, <u>11273</u> , <u>1642</u>
<i>D. melanogaster</i> modifiers genes using DIOPT orthologies not present in A	<u>10963</u> , <u>10992</u> , 51144, <u>7266</u> , <u>3920</u> , 5591, <u>1981</u> , <u>10294</u>
<i>D. melanogaster</i> modifiers genes using Ensembl orthologies not present in A	<u>5693</u>

Supplementary Table 3 (continuation). Ataxin-3 human interactors present in main databases, excluding the 151 in Sousa e Silva et al. [106], that are also expressed in reference SCA3 tissues, present in human polyQ database (N=252) in different EvoPPI3 predict interactomes databases

<i>D. rerio</i> predict polyQ_22 database using DIOPT and Ensembl orthologies not present in A	<u>7874</u> , <u>3735</u> , <u>1314</u> , 4627, 10121, 1937, 10575, 22948, <u>6217</u> , <u>5707</u> , 3376, <u>6194</u> , <u>10213</u> , 3191, <u>3192</u> , 10130, 1738, 2023, 7277, 6207, 6421, <u>2597</u> , 2316, 1743, 6203, 3921, 8661, 476, 8607, 4830, <u>1213</u> , <u>1615</u> , <u>5690</u> , 8402, 9861, 6188, 7167, 6202, <u>6301</u> , 6187, 11331, 5245, 10540, 9131, <u>6138</u> , 10787, 506, 26227, 5478, <u>6184</u> , <u>1938</u> , <u>10075</u> , <u>10606</u> , <u>9343</u> , 5717, 191, 10985, 1152, 1537, 3329, 3837, <u>4670</u> , <u>5718</u> , <u>3945</u> , 7203, <u>10399</u> , 5052, 6189, <u>23450</u> , <u>4673</u> , <u>5091</u> , 10935, 509, 6223, <u>498</u> , 6206, 5928, <u>372</u> , 539
<i>D. rerio</i> predict polyQ_22 database using DIOPT orthologies not present in A	7295, 4628, 3181, <u>3910</u> , 4841, <u>87</u> , 4967, 1832, <u>81</u> , 908, 7086, <u>1656</u> , <u>23191</u> , 84790, <u>3615</u> , 7001, 10549, 6210, <u>8773</u> , 3849, 6510
<i>D. rerio</i> predict polyQ_22 database using Ensembl orthologies not present in A	130162, 100996709, 51326

Supplementary Table 4. The distribution of the 81 ataxin-3 human interactors analyzed in Sousa e Silva et al. [106], that are expressed in reference SCA3 tissues and present in human polyQ database in different EvoPPI3

Predict interactomes databases	GeneID
B- <i>D. melanogaster</i> modifiers genes and <i>D. rerio</i> predict polyQ_22 databases	10808, 7317, 23464, 821, 3303, 3320, 7314, 6520, 119559, 3312, 10212, 7316, 3308, 5705, 22824, 203068
<i>D. melanogaster</i> modifiers genes using DIOPT and Ensembl orthologies not present in B	2033, 11284, 5710, 1387
<i>D. melanogaster</i> modifiers genes using DIOPT orthologies not present in B	7186, 3301, 292, 293
<i>D. melanogaster</i> modifiers genes using Ensembl orthologies not present in A	
<i>D. rerio</i> predict polyQ_22 database using DIOPT and Ensembl orthologies not present in B	7846, 5713, 5250, 6389, 6175, 7415, 90550, 84263, 10059, 47, 10999, 3420
<i>D. rerio</i> predict polyQ_22 database using DIOPT orthologies not present in B	391, 79751
<i>D. rerio</i> predict polyQ_22 database using Ensembl orthologies not present in B	130162, 100996709, 51326

Supplementary Table 5. The distribution of the 67 ataxin-3 human interactors analyzed in Sousa e Silva et al. [106], that are expressed in reference SCA3 tissues but not present in human polyQ database in different EvoPPI3 predict interactomes databases

Predict interactomes databases	GenelD
C- <i>D. melanogaster</i> modifiers genes and <i>D. rerio</i> predict polyQ_22 databases	8841, 824
<i>D. melanogaster</i> modifiers genes using DIOPT and Ensembl orthologies not present in C	834, 27005, 1460, 3300, 836, 387990, 7189
<i>D. melanogaster</i> modifiers genes using DIOPT orthologies not present in C	41, 7341, 7332
<i>D. melanogaster</i> modifiers genes using Ensembl orthologies not present in C	
<i>D. rerio</i> predict polyQ_22 database using DIOPT and Ensembl orthologies not present in C	6390, 6128, 6717, 396
<i>D. rerio</i> predict polyQ_22 database using DIOPT orthologies not present in C	1785, 8975, 373863, 7334, 4738, 310
<i>D. rerio</i> predict polyQ_22 database using Ensembl orthologies not present in C	

Supplementary Table 6. Proteins that present the same interaction behavior when using ataxin-3 I-TASSER predicted structure (IT-IT) and AlphaFold predicted 3D structures of predicted ataxin-3 interactors

GeneID	UniprotID	IT-AF dockings	IT-IT dockings
		JD - C-terminal interactions	JD - C-terminal interactions
41	P78348	4	38
267	Q9UKV5	59	36
293	P12236	60	11
310	P20073	28	16
391	P84095	47	21
396	P52565	39	26
821	P27824	11	44
824	P17655	36	26
834	P29466	6	30
871	P50454	15	39
1019	P11802	36	52
1026	P38936	-29	-12
1460	P67870	18	46
1468	Q9UBX3	37	31
2932	P49841	8	45
3105	P04439	8	47
3300	P25686	27	8
3303	P0DMV8	3	8
3308	P34932	11	49
3312	P11142	59	29
3420	O43837	22	44
4303	P98177	44	20
4697	O00483	15	16
4702	P51970	21	-2
4738	Q15843	24	26
4792	P25963	20	35
5071	O60260	49	32
5250	Q00325	25	4
5371	P29590	47	13
5631	P60891	20	32
5710	P55036	7	28
5886	P54725	9	44
6128	Q02878	24	36
6175	P05388	30	31
6383	P34741	45	39
6390	P21912	18	12
6520	P08195	18	30
6576	P53007	38	63
6717	P30626	34	29
7083	P04183	28	33
7157	P04637	19	27
7189	Q9Y4K3	-39	-16
7314	P0CG47	31	40

The protein-protein interactions (PPI) that cause Machado Joseph disease (SCA3)

Supplementary Table 6 (continuation). Proteins that present the same interaction behavior when using ataxin-3 I-TASSER predicted structure (IT-IT) and AlphaFold predicted 3D structures of predicted ataxin-3 interactors databases

GeneID	UniprotID	IT-AF dockings	IT-IT dockings
		JD - C-terminal interactions	JD - C-terminal interactions
7317	P22314	43	11
7326	P62253	25	27
7332	P68036	6	31
7334	P61088	31	33
7335	Q13404	39	28
7341	P63165	26	20
7846	Q71U36	33	31
8569	Q9BUB5	60	30
8841	O15379	36	27
8975	Q92995	39	25
9001	P54257	28	24
9524	Q9NZ01	28	30
9531	O95817	-10	-13
9894	Q9Y4R8	20	40
10013	Q9UBN7	14	35
10059	O00429	22	35
10212	O00148	41	3
10273	Q9UNE7	46	27
10797	P13995	17	20
10999	Q6P1M0	-16	-71
11284	Q96T60	2	28
23011	Q9UL25	27	34
23295	O60291	31	28
23464	O75600	38	34
23597	Q9Y305	40	31
23647	P53365	15	30
27338	Q16763	44	2
29979	Q9UMX0	35	47
51170	Q8NBQ5	41	29
54708	Q9NX47	14	37
56052	Q9BT22	41	18
57154	Q9HCE7	23	35
57159	Q9BYV2	9	25
65260	Q96BR5	2	42
78990	Q96DC9	31	21
79751	Q9H936	32	10
79888	Q8NF37	34	18
80262	Q9BSU1	7	49
81929	Q96EE3	26	33
83892	Q9H3F6	29	28
84263	Q6YN16	-40	-31
84447	Q86TM6	42	41
84676	Q969Q1	28	19
90550	Q8NE86	-53	-11
91942	Q8N183	15	35

Supplementary Table 6 (continuation). Proteins that present the same interaction behavior when using ataxin-3 I-TASSER predicted structure (IT-IT) and AlphaFold predicted 3D structures of predicted ataxin-3 interactors databases

GeneID	UniprotID	IT-AF dockings	IT-IT dockings
		JD - C-terminal interactions	JD - C-terminal interactions
112724	Q8NBN7	41	39
117584	Q8WZ73	40	38
119559	Q6P4A7	26	43
121536	Q6ZN18	34	20
200205	Q5T440	13	38
203068	P07437	11	32
337867	Q8NBM4	40	35
373863	Q8IYX4	48	12
387990	Q6UXN7	6	40
414919	Q6P1X6	43	54
493856	Q8N5K1	13	17
100287932	O14925	19	29

Supplementary Table 7. Interactors that mainly bind to the C-terminal region of wt ataxin-3 form using I-TASSER structures and AlphaFold interactors structure predictions

GeneID	Uniprot ID
498	P25705
5682	P25786
5902	P43487
4673	P55209
5700	P62191
10399	P63244
10992	Q13435
8678	Q14457
7431	P08670
5093	Q15365
3945	P07195

Supplementary Table 8. Interacting regions for each wt dataset and their correspondent interaction percentage (For the statistical analysis comparing the different interaction regions of each group, additional regions were added for both samples to have the same size. These additions do not reflect molecular signatures of the specific dataset and are marked in blue)

Interacting regions for the interactors from Chapter 3		Interacting regions for the interactors from Chapter 2	
Ataxin-3 wt I-TASSER residues	Interaction percentage	Ataxin-3 wt I-TASSER residues	Interaction percentage
42:GLU	24.10	42:GLU	50.00
43:GLU	50.60	43:GLU	56.00
44:GLU	33.73	44:GLU	56.00
45:ARG	2.41	45:ARG	31.00

Supplementary Table 8 (continuation). Interacting regions for each wt dataset and their correspondent interaction percentage (For the statistical analysis comparing the different interaction regions of each group, additional regions were added for both samples to have the same size. These additions do not reflect molecular signatures of the specific dataset and are marked in blue)

Interacting regions for the interactors from Chapter 3		Interacting regions for the interactors from Chapter 2	
<i>Ataxin-3 wt I-TASSER residues</i>	<i>Interaction percentage</i>	<i>Ataxin-3 wt I-TASSER residues</i>	<i>Interaction percentage</i>
46:MET	30.12	46:MET	48.00
47:ARG	60.24	47:ARG	70.00
48:MET	42.17	48:MET	59.00
49:ALA	10.84	49:ALA	25.00
50:GLU	43.37	50:GLU	50.00
51:GLY	43.37	51:GLY	60.00
52:GLY	42.17	52:GLY	59.00
53:VAL	39.76	53:VAL	59.00
54:THR	38.55	54:THR	62.00
55:SER	14.46	55:SER	38.00
56:GLU	26.51	56:GLU	46.00
57:ASP	40.96	57:ASP	62.00
85:LYS	67.47	85:LYS	69.00
86:VAL	63.86	86:VAL	62.00
87:TRP	67.47	87:TRP	66.00
88:GLY	69.88	88:GLY	69.00
89:LEU	66.27	89:LEU	60.00
90:GLU	62.65	90:GLU	70.00
91:LEU	19.28	91:LEU	28.00
92:ILE	48.19	92:ILE	57.00
99:TYR	59.04	99:TYR	66.00
100:GLN	19.28	100:GLN	33.00
101:ARG	40.96	101:ARG	60.00
102:LEU	67.47	102:LEU	70.00
103:ARG	80.72	103:ARG	85.00
104:ILE	48.19	104:ILE	64.00
105:ASP	57.83	105:ASP	68.00
106:PRO	15.66	106:PRO	25.00
107:ILE	69.88	107:ILE	59.00
108:ASN	85.54	108:ASN	82.00
109:GLU	69.88	109:GLU	65.00
110:ARG	83.13	110:ARG	78.00
125:LYS	72.29	125:LYS	63.00
126:LEU	8.43	126:LEU	15.00
127:GLY	74.70	127:GLY	54.00
128:LYS	77.11	128:LYS	71.00
129:GLN	33.73	129:GLN	53.00

Supplementary Table 8 (continuation). Interacting regions for each wt dataset and their correspondent interaction percentage (For the statistical analysis comparing the different interaction regions of each group, additional regions were added for both samples to have the same size. These additions do not reflect molecular signatures of the specific dataset and are marked in blue)

Interacting regions for the interactors from Chapter 3		Interacting regions for the interactors from Chapter 2	
<i>Ataxin-3 wt I-TASSER residues</i>	<i>Interaction percentage</i>	<i>Ataxin-3 wt I-TASSER residues</i>	<i>Interaction percentage</i>
166:LYS	67.47	166:LYS	66.00
167:GLY	66.27	167:GLY	68.00
168:ASP	86.75	168:ASP	77.00
169:LEU	80.72	169:LEU	62.00
170:PRO	78.31	170:PRO	56.00
171:ASP	80.72	171:ASP	62.00
175:ASP	66.27	175:ASP	61.00
176:GLN	45.78	176:GLN	46.00
177:LEU	24.10	177:LEU	42.00
178:LEU	72.29	178:LEU	72.00
179:GLN	91.57	179:GLN	81.00
180:MET	87.95	180:MET	73.00
181:ILE	48.19	181:ILE	70.00
182:ARG	16.87	182:ARG	52.00
183:VAL	32.53	183:VAL	56.00
184:GLN	78.31	184:GLN	79.00
185:GLN	30.12	185:GLN	47.00
186:MET	66.27	186:MET	71.00
187:HIS	22.89	187:HIS	46.00
188:ARG	85.54	188:ARG	71.00
189:PRO	37.35	189:PRO	34.00
190:LYS	65.06	190:LYS	55.00
191:LEU	57.83	191:LEU	56.00

Supplementary Table 9. Proteins with different supporting evidence for being ataxin-3 interactors

Database evidence for being ataxin-3 interactors	GeneID
Proteins present in Human polyQ database, <i>D. melanogaster</i> modifiers genes and <i>D. rerio</i> predict polyQ_22 databases	7531, 10574, 5901, 51520, 3309, 515, 5719, 226, 5689, 5686, 5702, 10694, 2058, 10383, 1072, 3326, 6191, 7514, 6950, 29927, 5706, 5709, 10594, 7407, 5708, 29968, 3313, 5230, 3190, 3310, 3305, 3315, 5315, 3304, 5701, 5700, 3276, 5682, 5685, 5704, 10576, 7431, 7184, 10382, 1778, 5093, 5687, 347688
Proteins present in Human polyQ database, <i>D. melanogaster</i> modifiers genes or <i>D. rerio</i> predict polyQ_22 databases	10527, 23197, 8678, 5902, 55660, 51639, 6117, 11273, 1642, 10963, 10992, 7266, 3920, 5591, 1981, 10294, 5693, 7874, 3735, 1314, 6217, 5707, 6194, 10213, 3192, 2597, 1213, 1615, 5690, 6301, 6138, 6184, 1938, 10075, 10606, 9343, 4670, 5718, 3945, 10399, 23450, 4673, 5091, 498, 372, 3910, 87, 81, 1656, 23191, 3615, 8773

Supplementary Table 10. Selection of the proteins that bind to the JD with different supporting evidence for being ataxin-3 interactors

	GeneID
Proteins that strongly bound to the Josephin domain	1213; 8773; 5718; 10213; 81; 10294; 55660; 10527; 226; 2597; 3315; 6184; 2058; 5091; 3615; 87; 3920; 1938; 1615; 5702; 6950; 10606; 5685; 1656; 6117; 5686; 5693; 10963; 5701; 5704; 372; 6301; 5690; 10694; 1314; 6138; 29927; 6217; 6191; 6194; 5901; 10576; 3192; 1981; 9343; 3735; 23450; 3910; 1642; 4670; 23191; 10075; 11273; 7874; 23197; 5707; 7266; 3276; 51520; 5719; 51639; 7514; 5709; 10382; 3326; 3304; 5315; 5689; 1072; 515; 7407; 3305; 3313; 5687; 7531; 5706; 10383; 347688; 10574; 29968; 3310; 5230; 5708
Proteins with enough evidence for being ataxin-3 interactors	7531; 10574; 5901; 51520; 3309; 515; 5719; 226; 5689; 5686; 5702; 10694; 2058; 10383; 1072; 3326; 6191; 7514; 6950; 29927; 5706; 5709; 10594; 7407; 5708; 29968; 3313; 5230; 3190; 3310; 3305; 3315; 5315; 3304; 5701; 5700; 3276; 5682; 5685; 5704; 10576; 7431; 7184; 10382; 1778; 5093; 5687; 347688
Proteins that strongly bound to the Josephin domain and enough evidence for being ataxin-3 interactors	10383; 7531; 10382; 5687; 5701; 10574; 5901; 347688; 1072; 51520; 10576; 3326; 3310; 515; 3305; 6191; 7514; 6950; 5719; 29927; 5706; 5709; 226; 3315; 5689; 3276; 7407; 5708; 5686; 29968; 5702; 5315; 5685; 10694; 5704; 3304; 2058; 3313; 5230

Supplementary Table 10 (continuation). Selection of the proteins that bind to the JD with different supporting evidence for being ataxin-3 interactors

	GeneID
Proteins that strongly bound to the Josephin domain and without enough evidence for being ataxin-3 interactors	7874 ; 3735 ; 10963 ; 10527 ; 23197 ; 1314 ; 5707 ; 6217 ; 6194 ; 10213 ; 3192 ; 3910 ; 3920 ; 2597 ; 87 ; 1981 ; 1213 ; 1615 ; 11273 ; 5690 ; 81 ; 6301 ; 5693 ; 6138 ; 6184 ; 1938 ; 1656 ; 10075 ; 10606 ; 9343 ; 4670 ; 5718 ; 55660 ; 7266 ; 23191 ; 51639 ; 3615 ; 6117 ; 23450 ; 5091 ; 8773 ; 372 ; 10294 ; 1642

Supplementary Table 11. Interacting regions and their correspondent interaction percentage for each group of proteins with more or less evidence for being ataxin-3 interactors (For the statistical analysis comparing the different interaction regions of each group, additional regions were added both samples to have the same size. These additions do not reflect molecular signatures of the specific dataset and are marked in blue)

Interactors with more evidence for being ataxin-3 interactors		Interactors with less evidence for being ataxin-3 interactors	
<i>Ataxin-3 wt I-TASSER residues</i>	<i>Interaction percentage</i>	<i>Ataxin-3 wt I-TASSER residues</i>	<i>Interaction percentage</i>
43:GLU	57.89	43:GLU	41.86
44:GLU	34.21	44:GLU	30.23
45:ARG	5.26	45:ARG	0.00
46:MET	34.21	46:MET	23.26
47:ARG	65.79	47:ARG	53.49
48:MET	50.00	48:MET	34.88
49:ALA	13.16	49:ALA	9.30
50:GLU	55.26	50:GLU	32.56
51:GLY	52.63	51:GLY	34.88
85:LYS	76.32	85:LYS	58.14
86:VAL	63.16	86:VAL	62.79
87:TRP	71.05	87:TRP	65.12
88:GLY	76.32	88:GLY	62.79
89:LEU	71.05	89:LEU	62.79
90:GLU	68.42	90:GLU	58.14
99:TYR	57.89	99:TYR	58.14
100:GLN	21.05	100:GLN	18.60
101:ARG	42.11	101:ARG	41.86
102:LEU	65.79	102:LEU	67.44
103:ARG	81.58	103:ARG	79.07
104:ILE	39.47	104:ILE	55.81
105:ASP	65.79	105:ASP	53.49
106:PRO	18.42	106:PRO	13.95
107:ILE	68.42	107:ILE	72.09
108:ASN	92.11	108:ASN	79.07
109:GLU	71.05	109:GLU	67.44
110:ARG	84.21	110:ARG	81.40

Supplementary Table 11 (continuation). Interacting regions and their correspondent interaction percentage for each group of proteins with more or less evidence for being ataxin-3 interactors (For the statistical analysis comparing the different interaction regions of each group, additional regions were added both samples to have the same size. These additions do not reflect molecular signatures of the specific dataset and are marked in blue)

Interactors with more evidence for being ataxin-3 interactors		Interactors with less evidence for being ataxin-3 interactors	
<i>Ataxin-3 wt I-TASSER residues</i>	<i>Interaction percentage</i>	<i>Ataxin-3 wt I-TASSER residues</i>	<i>Interaction percentage</i>
125:LYS	68.42	125:LYS	74.42
126:LEU	10.53	126:LEU	6.98
127:GLY	73.68	127:GLY	74.42
128:LYS	71.05	128:LYS	83.72
166:LYS	71.05	166:LYS	62.79
167:GLY	68.42	167:GLY	62.79
168:ASP	84.21	168:ASP	88.37
169:LEU	86.84	169:LEU	74.42
170:PRO	86.84	170:PRO	69.77
171:ASP	84.21	171:ASP	76.74
172:CYS	34.21	172:CYS	27.91
173:GLU	2.63	173:GLU	6.98
174:ALA	10.53	174:ALA	2.33
175:ASP	73.68	175:ASP	58.14
176:GLN	50.00	176:GLN	41.86
177:LEU	15.79	177:LEU	32.56
178:LEU	71.05	178:LEU	72.09
179:GLN	92.11	179:GLN	90.70
180:MET	94.74	180:MET	81.40
181:ILE	42.11	181:ILE	53.49
182:ARG	10.53	182:ARG	23.26
183:VAL	31.58	183:VAL	34.88
184:GLN	78.95	184:GLN	76.74
185:GLN	31.58	185:GLN	30.23
186:MET	71.05	186:MET	62.79
187:HIS	21.05	187:HIS	25.58
188:ARG	84.21	188:ARG	86.05
189:PRO	50.00	189:PRO	27.91
190:LYS	65.79	190:LYS	65.12
191:LEU	57.89	191:LEU	60.47

Supplementary Table 12. Differences of the number of interactions along expanded and wild type ataxin-3 forms when using I-TASSER and AlphaFold 3D protein structures considering the total number of interaction sites including the polyQ region and excluding it, as well as the percentage of the number of interaction residues with expanded form compared with wt form and the total number of interactions with wt ataxin-3

<i>GeneID</i>	<i>UniprotID</i>	<i>Total (exp-wt)</i>	<i>Total-polyQ (exp-wt)</i>	<i>Total (exp-wt)/Total (wt)(%)</i>	<i>Total-polyQ (exp-wt)/Total-polyQ (wt)(%)</i>
1213	J3KSQ2	-5	-5	-9,4	-9,4
8773	O00161	7	7	19,4	19,4
5718	O00232	6	6	14,0	14,0
10213	O00487	14	14	30,4	30,4
81	O43707	2	-	3,7	-
10294	O60884	5	-3	10,0	-6,0
55660	O75400	-1	1	-1,9	2,0
10527	O95373	-2	-2	-4,2	-4,2
226	P04075	-6	-6	-11,1	-11,1
2597	P04406	2	2	4,2	4,2
3315	P04792	-7	-7	-14,0	-14,0
6184	P04843	-3	-6	-5,9	-12,0
2058	P07814	-	-2	-	-3,5
5091	P11498	-10	-10	-16,1	-16,1
3615	P12268	-	-13	-	-26,5
87	P12814	9	9	20,9	20,9
3920	P13473	7	4	12,7	7,7
1938	P13639	5	5	10,4	10,4
1615	P14868	11	12	23,4	26,7
5702	P17980	14	6	29,8	12,8
6950	P17987	7	7	12,5	12,5
10606	P22234	-	-	-	-
5685	P25789	4	4	9,3	9,3
1656	P26196	-4	-4	-7,7	-7,7
6117	P27694	-5	-10	-9,3	-18,5
5686	P28066	4	4	10,0	10,0
5693	P28074	-	-7	-	-13,0
10963	P31948	12	7	27,9	16,3
5701	P35998	6	6	12,2	12,2
5704	P43686	16	16	36,4	36,4
372	P48444	-3	-3	-5,9	-5,9
6301	P49591	-4	-4	-7,3	-7,3
5690	P49721	8	8	18,2	18,2
10694	P50990	-4	-4	-6,3	-6,3
1314	P53621	12	12	23,1	23,1
6138	P61313	20	17	43,5	37,0
29927	P61619	5	5	10,0	10,0
6217	P62249	-4	-4	-8,7	-8,7
6191	P62701	8	8	17,8	17,8
6194	P62753	10	-1	25,6	-2,6
5901	P62826	1	1	2,2	2,2

Supplementary Table 12 (continuation). Differences of the number of interactions along expanded and wild type ataxin-3 forms when using I-TASSER and AlphaFold 3D protein structures considering the total number of interaction sites including the polyQ region and excluding it, as well as the percentage of the number of interaction residues with expanded form compared with wt form and the total number of interactions with wt ataxin-3

<i>GeneID</i>	<i>UniprotID</i>	<i>Total (exp-wt)</i>	<i>Total-polyQ (exp-wt)</i>	<i>Total (exp- wt)/Total (wt)(%)</i>	<i>Total-polyQ (exp- wt)/Total-polyQ(wt)(%)</i>
10576	P78371	-15	-24	-24,2	-38,7
3192	Q00839	2	1	4,1	2,0
1981	Q04637	-	-	-	-
9343	Q15029	-5	-5	-9,8	-9,8
23450	Q15393	10	10	19,2	19,2
3910	Q16363	6	6	9,1	9,1
1642	Q16531	5	5	9,8	9,8
4670	Q59ES8	3	3	6,1	6,1
23191	Q7L576	-5	-26	-8,6	-44,8
10075	Q7Z6Z7	9	6	20,0	13,3
11273	Q8WWM7	-17	-17	-24,6	-24,6
7874	Q93009	2	2	3,4	3,4
23197	Q96CS3	-8	-8	-15,1	-15,1
5707	Q99460	17	12	37,8	26,7
7266	Q99615	-2	-2	-4,2	-4,2
3276	Q99873	-	-	-	-
51520	Q9P2J5	-13	-21	-25,0	-40,4
5719	Q9UNM6	1	1	2,2	2,2
51639	Q9Y3B4	3	3	7,0	7,0
7514	O14980	16	1	20,5	1,3
5709	O43242	-8	-8	-12,3	-12,3
10382	P04350	-6	-6	-11,1	-11,1
3326	P08238	9	-2	20,5	-4,5
3304	P0DMV9	-6	-21	-9,2	-32,8
5315	P14618	1	1	1,9	1,9
5689	P20618	2	7	3,6	14,0
1072	P23528	-5	-5	-10,6	-10,6
515	P24539	26	26	54,2	54,2
7407	P26640	-6	-6	-12,0	-12,0
3305	P34931	2	-2	3,4	-3,4
3313	P38646	-5	-5	-10,2	-10,2
5687	P60900	4	4	8,0	8,0
7531	P62258	-2	-2	-3,6	-3,6
5706	P62333	7	-1	11,9	-1,7
10383	P68371	6	6	14,0	14,0
347688	Q3ZCM7	-	-	-	-
10574	Q99832	5	4	8,8	7,7
29968	Q9Y617	-16	-16	-31,4	-31,4
3310	P17066	-23	-41	-28,4	-50,6
5230	P00558	-14	-22	-20,3	-32,8
5708	Q13200	3	-12	5,0	-21,1
3735	Q15046	5	1	10,6	2,1

